PREDOMINANCE OF SIX HEXANUCLEOTIDE RECODING SIGNALS 3' OF READ-THROUGH STOP CODONS

By

LANCE RORK HARRELL

Bachelor of Arts

Concordia College

Moorhead, Minnesota

1993

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE

August, 2002

PREDOMINANCE OF SIX HEXANUCLEOTIDE

RECODING SIGNALS 3' OF READ-THROUGH

STOP CODONS

Thesis Approved:

Ulrich Melcher

Thesis Advisor

Richard C Essenberg

Hani S Botrous

Kay Scheets

Timothy J. Pettitt

Dean of the Graduate College

PREFACE

Molecular genetic studies of read-through stop codons in selected viruses have revealed that the sequences surrounding the stop codon reduce the fidelity of translation termination. To determine to what extent the sequence preferences revealed in those studies apply to read-through events in other viruses, I examined the distribution of residues surrounding read-through stop codons. I compared sequence segments from every unique viral sequence annotated as having a read-through stop codon. Inspection of the codons immediately following read-through stop codons led to identification of six contexts that accounted for almost 90% of the sequences examined. Chi-square analysis of sequence variability in these contexts demonstrated that, for five of the six, the sequences immediately 3' of read-through stop codons are conserved within each group. In contrast, for the non-read-through stop codon of the read-through gene, no such conservation occurred. The relative efficiency of these read-through contexts in mammalian tissue culture cells has been determined using a dual luciferase fusion reporter. These observations support the hypotheses that translation read-through is a signaled event and that a limited number of distinct sequences provide this signal.

ACKNOWLEDGEMENTS

I would like to acknowledge the members of my committee — Dr. Richard Essenberg, Dr. Marie Petracek and Dr. Kay Scheets — here for taking the time to provide me with their insight and constructive criticism on this thesis. I would also like to acknowledge Dr. John Atkins for his input and aid, in not only the luciferase portion of this research, but also in the formation of the resulting manuscript. I would especially like to thank my advisor, Dr. Ulrich Melcher, for providing an ideal environment for developing both the physical and intellectual skills required to complete my degree. I would also like to thank Ulrich for his understanding and for his friendship. I sincerely hope that we remain in contact after my departure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

Å	Angstroms
Ave % RT	average percent read-through
BYDV-PAV	*Barley yellow dwarf virus PAV*
COS	simian fibroblast CV-1 cells
CP	coat protein
DMEM	Dulbecco's modified eagle medium
DMEM-10%-FBS	DMEM containing 10% fetal bovine serum
DMEM-SF	serum free DMEM
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
FS	frame shifting
Hek 293	human embryo kidney number 293 cells
IRES	internal ribosome entry site
ITAF	IRES trans-acting factor
kb	kilobases
kDa	kilodaltons
ml	milliliters
MuLV	*Murine leukemia virus*
nt	nucleotides
ng	nanogram
RF	release factor
RLU	relative light units

RNA	ribonucleic acid
rRNA	ribosomal RNA
tRNA	transfer RNA
SBP2	SECIS binding protein
SECIS	selenocysteine insertion sequence
S	Svedberg
TMV	*Tobacco mosaic virus*
TRV	*Tobacco rattle virus*
µl	microliters
xg	times gravity

CHAPTER I: INTRODUCTION

Translation

All genetically inheritable information discovered thus far is transmitted in the form of nucleic acid, either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). Containing only the four bases, Adenine, Guanidine, Cytosine and Thymine (DNA)/ Uracil (RNA), along with various modifications, these nucleic acids provide a non-overlapping, comma free, degenerate, triplet code which serves as the blueprint for protein construction. Some of the major work that led to these conclusions came from Crick and Brenner (1). Working with *Bacteriophage T4*, they created proflavin-induced mutations in T4's *rIIB* cistron that restricted the bacteriophage's ability to grow in a restrictive host. They found that the mutations were of two types, which they termed + or -, corresponding to either an addition or deletion respectively. Crick and Brenner noticed that when double mutants were constructed, + + and - - mutants still displayed the mutant phenotype, while heterogeneous mutations such as + - or - + would display the wild-type phenotype. They concluded that the genetic code is read from a fixed starting point, and is hence, comma free. They extended their work by creating triple mutants, and found that only + + + or - - - combinations would restore the wild-type phenotype. This data supported the idea that the genetic code was indeed triplet in nature; and that the code was non-overlapping because other heterogeneous triple mutants disrupted the translation of the remainder of the gene. Around the same time, Yanofsky demonstrated that genes and the polypeptides that are derived from them are co-linear (2). This was accomplished using a combination of transductional mapping and fingerprinting to show that the

location of mutations in the gene for tryptophan synthase corresponded to amino acid changes at a certain position in the polypeptide.

With this information in hand, the next step was to determine which nucleotide triplet, also termed a codon, coded for what amino acid. The genetic code was deciphered using two in vitro experimental systems, both pioneered by Nirenberg and developed by Khorana in the early 1960s. The earlier of the two systems involved creating a homogeneous RNA molecule, such as a poly U, or a repeating triplet, such as ...AUUAUU..., and determining which amino acid was incorporated into the polypeptide (3). This system had obvious limitations, so a second system was developed which used a synthesized tri-nucleotide to bind a charged tRNA in the A-site of the ribosome. The ribosomes, with bound tRNA, could then be retained by a nitrocellulose membrane while the free tRNA could be washed away. The system used a mixture of charged tRNAs, but with only one amino acid radioactively labeled. This allowed for the screening of different amino acids against one codon, making it possible to assign most of the possible nucleotide triplet combinations to the appropriate amino acid. Using this system it was discovered that more than one nucleotide triplet coded for a particular amino acid, that is, the genetic code is degenerate (4,5). It was also discovered that three of the 64 possible tri-nucleotides did not code for an amino acid. Further work revealed that these codons, UAA, UAG and UGA, signaled the ribosome to terminate translation, and were therefore termed stop codons.

Since DNA does not directly code for polypeptides, but rather uses an intermediate, messenger RNA (mRNA), it is on the mRNA where the nucleic acid code is translated into one consisting of amino acids. However, since an amino acid cannot

directly recognize its corresponding codon, Crick proposed the "adaptor hypothesis" which states that translation occurs via a molecule which can recognize a specific codon and also has attached to itself the appropriate corresponding amino acid. Zamecnik and Hoagland later advanced this hypothesis by discovering that radioactively labeled amino acids became transiently bound to the low molecular mass fraction of RNA, which they called soluble RNA (sRNA). Further investigation proved that the sRNA molecules, which eventually became known as transfer RNAs (tRNA), were indeed the adapters Crick envisioned.

Further investigation of tRNAs revealed that they vary in length from about 60 to 95 nucleotides (nt), with an average length of 76 nt. The internal base pairing that occurs in the molecule is typically described in terms of a cloverleaf structure containing a stem and three stem loops (6). Moving along a typical tRNA, starting at the 5' end of the molecule, the stem, or acceptor stem, contains the 5' terminal phosphate group, is typically seven base pairs (bp) in length, and may contain non-Watson-Crick base pairs. The first stem loop, the D arm, consists of a three or four bp stem ending in a loop that frequently contains dihydrouridine. The anticodon arm, the next stem loop encountered, is a combination of a five bp stem and a loop that contains the anticodon. The last leaf of the clover is the T arm, which is made up of a five bp stem ending in a loop that usually contains the sequence TΨC, where Ψ is a pseudouridine. The 3' end of the molecule base pairs with the acceptor stem, forming a single stranded CCA overhang with a free hydroxyl group, thus completing the cloverleaf. In addition to these standard secondary structure features, there may also be a variable arm between the anticodon arm and the T arm, which has been known to be anywhere from three to 21 nt in length.

The primary sequence of a tRNA is heavily modified, with up to 25% of the nucleotides changed in some fashion. Of all the tRNAs examined, a total of 80 bases at 60 different positions have been identified as targets for modification. The tRNA folds into an L-shaped tertiary structure (7). The acceptor stem and T arm make up one leg of the "L", by folding into a continuous A-RNA-like double helix. The other leg is formed in a similar fashion by the D arm and the anticodon arm. The L shape is important as it makes the molecule only 20 to 25 Å in width, allowing two tRNAs to bind close to one another during translation.

tRNAs were the adaptors that Crick envisioned, but how do they become associated or "charged" with the proper amino acid? As it turns out, each amino acid has its own enzyme that oversees charging. These enzymes, known as aminoacyl-tRNA synthetases, catalyze an ATP driven reaction that attaches the proper amino acid to the corresponding tRNA. Two classes of aminoacyl-tRNA synthetases exist (8). The first class attaches the amino acid to the 2' hydroxyl of the tRNA acceptor stem, and requires recognition of the anticodon for charging. This class consists of proteins containing two parts of the Rossmann fold, the HIGH sequence (His-Ile-Gly-His) and the KMSKS sequence (Lys-Met-Ser-Lys-Ser), that are known to be important in ATP binding. The second class of aminoacyl-tRNA synthetases attaches the amino acid to the 3' hydroxyl of the acceptor stem and does not require anticodon recognition for charging. These proteins all share a core catalytic domain that consists of a seven stranded antiparallel β-sheet with three flanking helices. Although the first class utilizes the anticodon for recognition, while the second class does not, both classes use regions of the acceptor stem, variable arm, and D arm to ensure proper charging.

Once the tRNA is charged, the next step in creating a polypeptide is the decoding of the mRNA by the tRNA and the connection of one amino acid to another. This step occurs with the employment of a large multi-subunit protein and RNA complex called the ribosome. Prokaryotic and eukaryotic ribosomes are similar, however, more is currently known about the structure/function relationships of the smaller prokaryotic ribosome, so that is what I will outline here.

The *Escherichia coli* (*E.coli*) ribosome consists of two subunits. The 30S subunit binds the mRNA guided by the translation initiation region of the mRNA (9,10) and other factors. The 50S subunit then binds to the 30S/mRNA complex, enclosing the mRNA and forming three pockets in the ribosome, A, P and E. Then, as the ribosome ratchets along the mRNA, codon by codon, during elongation (discussed below), a codon first enters the A-site where tRNA recognition of the mRNA occurs. The codon then enters the P-site after transpeptidation and translocation (discussed below). The E-site is occupied by the tRNA that was displaced from the P-site during the previous cycle of elongation after transferring the peptidyl residue to the aminoacyl-tRNA in the A-site.

The two ribosome subunits are themselves made up of several components. The 30S subunit contains a 16S ribosomal RNA (rRNA) and 21 polypeptides, while the 50S subunit contains a 23S and a 5S rRNA, as well as 31 polypeptides. There is much more information on the structure and function of ribosomal components then would be practical to present here. Some of the pertinent relationships are the site of mRNA binding, which occurs at the 3' end of the 16S rRNA (11), the tRNA anticodon binding sites, which occur in the 30S subunit's "cleft" region (12-14), and the peptidyl transferase site, which occurs in the "valley" between the 50S subunit's other two protuberances.

Ribosomes are either free in the cytoplasm of the cell or can be bound to the endoplasmic reticulum (ER) in eukaryotes. If the ribosome is on the ER, it binds the ER membrane near the exit tunnel on the larger subunit.

The mRNA, tRNA and ribosome are the three major players in the translational process. The process translates the mRNA in the 5' to 3' direction, and produces the polypeptide in the N-terminal to C-terminal direction (15). Translation can be thought of as a three stage process starting with chain initiation, progressing with chain elongation, and ending with chain termination. Since elongation and termination are the two most relevant stages of translation in regards to the research that will be described later, chain initiation will not be discussed here (16).

Elongation of the polypeptide chain is a cyclical mechanism that can process up to 40 amino acid residues per second. The cycle requires additional, non-ribosomal proteins termed elongation factors (EFs). Again, as is the case for ribosomal components, prokaryotes and eukaryotes employ different but similar elongation factors. So again, for simplicity's sake, I will discuss the cycle using the prokaryotic system. The elongation cycle is best described in three phases: binding, transpeptidation, and translocation. It is important to remember that at this phase in translation, the ribosome has bound the mRNA and is somewhere downstream of the AUG start codon.

During the binding phase of the elongation cycle, an EF-Tu-GTP complex combines with an aminoacyl-tRNA. This complex then enters the A-site of the ribosome where the complex is either accepted or rejected based on the codon/anticodon interactions of the mRNA with the tRNA. If the tRNA is accepted, it is bound to the codon in a GTP hydrolysis dependant manner that releases EF-Tu/GDP and inorganic

phosphate. With the aminoacyl-tRNA now bound in the A-site of the ribosome, the next phase of the elongation cycle, transpeptidation, occurs as the peptide bond is formed between the tRNA-linked polypeptide in the P-site and the tRNA-linked amino acid in the A-site. This occurs via a nucleophillic displacement reaction when the amino group of the 3'-linked aminoacyl-tRNA in the A-site displaces the tRNA portion of the peptidyl-tRNA in the P-site, thereby transferring the polypeptide to the A-site tRNA. It is interesting to note that the displacement does not require any outside energy input as the bond between the nascent polypeptide and the P-site tRNA is a high energy bond. The last phase of the elongation cycle is the translocation of the A-site tRNA linked polypeptide into the P-site of the ribosome. The process requires the assistance of EF-G, which hydrolyses GTP causing the expulsion of the P-site tRNA into the E-site and the movement of the ribosome one codon further down the mRNA.

Recall from the earlier discussion of the genetic code, that three codons signal for the ribosome to terminate translation: UAA, UAG and UGA. Chain termination proceeds in a very similar manner as does one cycle of chain elongation. However, in termination, there are no tRNAs that bind to the stop codons. Instead, specialized proteins called release factors (RFs) fulfill the role. When one of the stop codons is encountered in the A-site of a prokaryotic ribosome, either RF-1 (for UAA and UAG) or RF-2 (for UAA and UGA), both of which are tRNA shaped, interact with the stop codon. The binding of RF-1 or RF-2 induces the ribosomal peptidyl transferase to transfer the polypeptide to H_2O rather than to the aminoacyl-tRNA that would otherwise be present in the A-site. This transfer therefore releases the polypeptide from the translational machinery, and terminates translation of the mRNA by that particular ribosome. RF-3/GTP then binds to

and stimulates the release of RF-1 or RF-2 from the ribosome (17,18). In eukaryotes, the situation is the same, except that a protein dimer (eRF-1/eRF-3) takes the place of RF-1 and RF-2 by binding all three types of stop codon (19).

Translational Exceptions

Our incomplete understanding of protein synthesis results in occasions when translation behaves in ways that appear contrary to the existing model. The existence of any such deviation seems at first to be counter-productive, but when viewed as a strategy for translational control, the purpose of such a mechanism becomes more apparent. Rather than a definition of translational control as referring to the amount of a protein produced, an alternative definition can be constructed which refers to the regulation of the structure of the encoded protein. This type of control, called recoding, occurs by occasionally allowing for an additional domain to be translated, essentially producing two protein products from the same mRNA. Recoding occurs when a sequence in the mRNA causes the ribosome to act in a manner different from the standard rules, allowing the ribosome to ignore the stop signal and continue elongating the polypeptide.

Three forms of recoding exist, the first of which is frameshifting (FS). Both types of FS, -1 and +1, involve changing the reading frame during translation in such a way that the component of a protein produced before the stop codon and the component produced after the stop codon are translated from different reading frames. –1 FS is widely used by retroviruses and retrotransposons, as well as by other viruses, to express a defined ratio of a long fusion protein such as a *gag-pol* fusion protein (20). –1 FS results in the shift of the reading frame by one nucleotide toward the 5' end of the mRNA and, except for a few exceptions, is implemented via a single mechanism. The mechanism,

known as simultaneous slippage, requires the presence of the slippery hexamer sequence (X-XXY-YYZ) upstream of the terminator. X represents one of the four nucleotides, Y represents a weakly base pairing nucleotide and Z represents a nucleotide that is species specific. The frameshifting proceeds when two tRNAs decoding XXY-YYZ slide backward on the mRNA one nucleotide to XXX-YYY (21,22). In addition to the slippery hexamer, efficient frameshifting usually requires the mRNA to form some sort of secondary structure, either a stem-loop or pseudoknot, 3' of the stop codon (23-25).

The second type of frameshifting occurs by shifting the reading frame one nucleotide toward the 3' end of the mRNA, and is hence termed +1 FS. This type of recoding event is less common than –1 FS, and occurs at a different stage of the elongation cycle. For +1 FS to take place, the last 0 frame codon must code for a low abundance aminoacyl-tRNA (26), or a poorly recognized stop codon with the proper 3' context (27). The low abundance of the tRNA in the former, or the slow recognition in the latter, causes the ribosome to pause, allowing the tRNA that recognizes the +1 codon to enter the A-site and promote the shift.

The second form of recoding, ribosome hopping, has been best characterized in gene *60* mRNA of bacteriophage T4 (28), so the discussion of hopping will take place in that context. When the ribosome encounters a UAG terminator at codon 47 in gene *60*, about 50% of the time the peptidyl-tRNA^{gly} dissociates from the 46th codon while the other 50% remain bound to the ribosome. The ribosome then "hops" down the mRNA where the tRNA rebinds to an identical codon 48 nt downstream. The process is broken down into three phases: takeoff, where the tRNA releases the mRNA; scanning, where the ribosome searches for the downstream gly codon; and landing, where the tRNA

rebinds to the mRNA and translation resumes. Hopping requires three cis-acting elements; the in-frame UAG at codon 47, an upstream primary sequence of the nascent protein, and a downstream stem-loop structure (29).

The last, and least understood, form of recoding is that of read-through and occurs when the ribosome misreads a nonsense codon as sense. Most read-through occurs by the mis-incorporation of an aminoacyl-tRNA for the terminator; the most common being glutamine for UAA or UAG and tryptophan for UGA (30). Programmed read-through of a terminator is insured through the use of signal sequences built into the mRNA sequence. The simplest signals may consist of only a short sequence 3' of the stop codon. This is demonstrated by the read-through of the *Sindbis virus nsP4* gene being facilitated by a single C immediately following the UGA terminator (31), or the CAR-YYA sequence found to be important for the in vivo read-through of the UAG of *Tobacco mosaic virus* (TMV) RNA (32) and the in vitro read-through of UAG, UAA and UGA in a TMV specific context (32,33). The more complex signals have been known to involve a proximal sequence and secondary structure as well as a distal sequence as shown by in vivo and in vitro deletion studies for the UAG terminator in *Barley yellow dwarf virus PAV* (BYDV-PAV) RNA (34).

How these different sequences signal read-through is still the topic of current investigation. Evidence suggests that in prokaryotes there is an interaction involving the stop codon and two parts of the 16S rRNA (C1054 of helix 34 and the A-site portion of helix 44) that form a structurally unique surface recognized by release factors (35). There is also evidence that a similar mechanism is at work in eukaryotes because of the conservation of C1054 in yeast 18S rRNA (36). Also, the strength of the terminator's

interaction may be due in part to the nucleotide sequence surrounding the stop codon (37). This is supported by site-directed crosslinking experiments showing that nucleotide positions +1 (38) and +4 to +6 (39) can be crosslinked to RF. Therefore, it could be possible that the read-through signals reduce the efficiency of RF binding, causing the ribosome to pause and thereby allowing noncognate decoding of the terminator. However, more complex interactions may also occur between the sequence or its secondary structure and a ribosome-associated factor. Ribosomal proteins, rRNA and release factors are all possible candidates for the interaction, and work continues to determine if any or all may play a part.

Selenocysteine, also referred to as the 21st amino acid (40), forms the basis of a second form of read-through. Unlike the read-through systems just discussed, in selenocysteine incorporation, a *specific* aminoacyl-tRNA is produced that decodes the terminator as a sense codon. This type of recoding is employed by several bacteria (41), as well as a variety of other organisms ranging from protozoa to vertebrates (42). Although this event is present in both prokaryotes and eukaryotes, the mechanism used by each is different. Prokaryotes require a highly conserved stem-loop structure at a specific position immediately 3' of the terminator (43). The stem-loop, also called the selenocysteine insertion sequence (SECIS), binds the selenocysteine-tRNA via a special elongation factor, SEL B, and waits for the ribosome to encounter the terminator.

In contrast to prokaryotes, eukaryotes all have their SECIS in the 3'UTR of the transcript (44). This immediately presents the question of how an element so distant from its site of action can function. There is also an additional element to consider; the single eukaryotic SECIS element conveys the ability to recode *all* of the UGA terminators in a

transcript. Two models have been proposed to account for both of these observation, both of which consist of the same list of players: the SECIS element; the SECIS binding protein (SBP2); eEFsec, which binds both SBP2 and $tRNA^{sel}$; and the $tRNA^{sel}$.

In the first model, SBP2 binds the SECIS element and stimulates the binding of the $tRNA^{sel}$/eEFsec complex. This complex then reaches back and inserts the $tRNA^{sel}$ when the ribosome reaches the UGA terminator (45). The second model suggests that the complete SECIS/SBP2/eEFsel/$tRNA^{sel}$ complex forms, then attaches to the ribosome at the beginning of translation and "rides" along until a UGA stop is encountered (46). While both models account for the observed phenomenon, more investigation is needed to determine their validity.

Viruses

A virus is an obligate parasite that consists of a nucleic acid genome that codes for, among other proteins, the capsid protein(s) in which it is encased. The nucleic acid component may be either RNA or DNA, and can be single stranded or double stranded. The capsid can be made from a single type, or multiple types of coat proteins (CP) and may be covered with a membrane envelope. Viruses utilize the cellular machinery of their host to replicate, making them useful tools for gaining insight into cellular processes such as gene expression and translation.

A typical animal virus infection cycle begins with the virus recognizing specific features of the host cell surface that will allow the virus to introduce its genome into the cell. This introduction can either be by injecting the nucleic acid through the plasma membrane or by stimulating the cell to engulf the entire virus particle (47). By contrast, a plant virus must take advantage of a break in the plant cell wall to gain access to the

plasma membrane, where the virus can apparently enter the cell without interacting with specific receptors (48). Once the viral genome has entered the cell, the viral genes must be decoded and translated for the virus to replicate and spread. Recall that there are different types of viral genomes, so the cellular machinery that an individual virus type needs to utilize varies. In the last stage of infection, new virus particles must be formed. New virus genomes are usually produced concurrently with a second stage of viral gene expression that ensures the presence of capsid protein(s). The genomes, whether associated with the capsid protein in the form of virions or not, leave the host cell, either by virus shedding, cell lysis, or in the case of plant viruses, plasmodesmata.

In utilizing the host cell machinery, viruses that regulate their gene expression have the problem that eukaryotic ribosomes normally cannot initiate translation of an open reading frame following translation of the ORF immediately preceding it. Viruses utilize several strategies to overcome this limitation, the simplest being to have a segmented genome.

Cowpea mosaic virus produces and encapsidates two RNA molecules in two separate isohedral particles (49). One 5.9 kb RNA produces a polyprotein that is cleaved into a 24 kDa protease, 4 kDa VPg, 110 kDa replicase and a 32 kDa processing protein. The 3.5 kb RNA produces a polyprotein that is cleaved into the 42 and 24 kDa capsid proteins as well as the set of proteins needed for virus movement. The caveat to this type of strategy is that for a successful infection the host cell must be infected by both types of particle. This problem is exemplified by *Brome mosaic virus* which encapsidates not two, but three different RNAs, one RNA each encoding either the replicase, capping enzyme or movement protein and coat protein.

Another strategy involves the generation of sub-genomic RNAs (50). Expression from the full length, (+) sense RNA results in the production of a protein from a 5' ORF. This protein then helps direct the generation of other, smaller mRNAs from the full (–) sense RNA by transcription, allowing the production of a different set of proteins than those encoded by the original full-length transcript.

A special type of this strategy, called 3' discontinuous extension of negative strands, is employed by *Coronaviruses* (51). The translation of the genomic RNA produces a replicase protein. The replicase then generates a full-length (–) sense RNA using the virion (+) sense RNA as a template. Starting from the 3' end of the (–) sense RNA, smaller (+) sense RNAs are produced by the replicase which transcribes an identical leader sequence, then jumps to specific points on the(–) strand template before continuing.

The process known as internal entry of ribosomes is another method utilized to control gene expression in viruses as diverse as *Flaviviruses*, *Retroviruses* and *Picornoviruses*. This strategy requires that the sequence of the mRNA contain an internal ribosome entry site (IRES) which forms a pseudoknot structure (52,53) that interacts with a combination of several IRES trans-acting factors (ITAFs) that mediate IRES-dependent translation initiation (54-56). Some viruses, such as *Picornoviruses*, use this approach of controlling gene expression in combination with encoding a protease that inactivates eukaryotic initiation factor 4. The inactivation effectively shuts down cap dependent translation and better allows the virus to "hijack" the cell's translational capabilities (57).

Simian virus 40 (SV40) is an example of a virus that uses another strategy known as alternative splicing. In the case of SV40, alternative splicing produces two proteins,

the small and large T antigen, from a single RNA. The small T antigen is produced through the "normal" translational process. The large T antigen is produced by the removal of the stop codon region of the RNA, between the end of the small T antigen and the beginning of the large T antigen, and splicing together the two coding regions.

The final strategy known to be employed by viruses to differentially express their genes is that of recoding, which was covered in the Translational Exceptions section of Chapter I. These strategies are not mutually exclusive, and many viruses use more than one to fulfill their needs.

New Opportunities

The continuing development of two tools at the disposal of the molecular biologist, databases and reporter genes, have made the work presented here possible. Databases are not new. Paper based bibliographic databases such as Chemical Abstracts and Biological Abstracts have been around for almost a century. However, computer technology has facilitated an increase in the feasibility and usability of computer based factual databases. Factual databases, such as GenBank, the European Molecular Biology Laboratory (EMBL) and DNA Database of Japan (DDBJ) for nucleic acids, are expanding rapidly thanks to development of DNA sequencing technology. The large number of entries in these factual databases have allowed for the creation of a second generation database, the knowledge base. Unlike bibliographic or factual databases, which are designed only for information retrieval, knowledge bases are intended to generate new information for existing data. One example of such a knowledge base is the Mouse Aging knowledge base organized by Richard Miller and David Burke at the University of Michigan Medical School (58). The foundational data for the database is

generated by first harvesting a mouse population and separating the individuals various parts (heart, bone, muscle, brain, etc). Each part is then sent to a separate researcher specializing in that tissue. The data generated by the different specialists is entered into the database. This phenotypic and genotypic information can be compared to draw conclusions about how specific genotypic changes incur phenotypic changes associated with aging.

Knowledge bases, however, are not the ultimate expression of this trend. In fact, a third generation "database", referred to as the virtual cell, has been proposed. The virtual cell would combine the overall biological knowledge gained from molecular and cellular biology, biochemistry, genetics, etc. with complete genomic sequences and the information obtained from perturbation (environmental change, gene disruption, etc.) of living cells, in an attempt to use this information to determine the basic principles and complex interactions that take place within a cell (59). Obviously, this is an enormous task, not only in attempting to create a computerized representation of a cell, but also in the organization of the massive amount of information the virtual cell would need to access.

The second tool, reporter genes, allows for the detection of a previously undetectable cellular event by linking that event with another that can be detected. Some popular methods include β-complementation, of which the β-galactosidase system is a common example. Another method involves the direct linking of a gene product with another product that can be easily detected, such as fusions with green fluorescent protein (60).

The research presented below utilized a reporter system based on two different luciferase genes, one from *Renilla* and one from *Photinus* (61). The genes are placed in a vector such that small sequences can be inserted between them. This system is an excellent tool to assess translation termination signaling, since the efficiency of the inserted termination sequences can be tested by assaying for the activity of downstream luciferase enzyme.

While several experiments have been conducted to find the signaling sequences important for recoding using this or other methods, each project has focused on only one, or at most, a few viruses. One such study identified the signaling sequence for read-through of the UGA terminator between the *nsP3* and *nsP4* genes of *Sindbis virus* (31). By subcloning the 612 bp region surrounding the leaky terminator into an SP6 transcription vector, a fusion with the *Sindbis virus* capsid protein was created. The ability of this construct to generate the read-through fusion protein, approximately 10% of the total protein produced, demonstrated that the read-through signal was contained within the 612 bp region. To further define the read-through signal, a second construct was made placing the three codons upstream and the four codons downstream of the leaky stop codon between a truncated chloramphenicol acetyltransferase (CAT) gene and the 4D4 epitope of *Rift valley fever virus*. This construct allowed for the identification of the translation products by both size and immunoprecipitation. The shortened insert still maintained the ability to facilitate read-through at approximately 10% when translated in rabbit reticulocyte lysate. Additional definition of the signal was performed by inserting just the leaky UGA and the 3' proximal CUA codon into two contexts, the *Sindbis* capsid protein system mentioned above and the M gene system from mice. In both cases the

insert still signaled for read-through of the stop at approximately 10%. With the signal now identified as being within the 3' proximal triplet, mutational analysis of this triplet showed that only the 3' proximal nucleotide, a cytosine, was sufficient to signal read-through for *Sindbis virus*.

The read-through signal for the UAG terminator between the helicase and replicase genes in TMV was characterized in a similar manner (32). The location of the read-through signal was narrowed down by inserting the UAG, complete with eight 5' codons and six 3' codons, into a glucuronidase expression vector which was then expressed in tobacco protoplasts (62). The insert was able to signal read-through at wild-type levels of approximately 5% demonstrating that the signal was within this region. Next, a series of 5' and 3' codon deletions were made to narrow the location of the signal further. The deletion experiments showed that none of the 5' codons were part of the signal, but if either of the two most proximal 3' codons were removed, read-through decreased 20 fold. These two codons were examined in more detail by performing a mutational analysis which revealed that the read-through signal for the UAG of TMV consists of the consensus sequence CAR-YYA for the two codons immediately following the terminator.

Both of the above studies are cases where the read-through signal can be traced back to a relatively simple sequence in the mRNA. In the following studies however, the signal appears to be more complex in nature. The determination of the read-through signal for the UAG between the *gag* and *pol* genes of *Murine leukemia virus* (MuLV) was characterized by creating a series of constructs containing the two upstream codons and either 2, 5, 10, 15 or 19 downstream codons inserted between a CAT and lacZ genes

of a transcription vector (63). In vitro transcription and translation in reticulocyte lysate followed by immunoprecipitation with a CAT antibody revealed that only the construct with the nineteen downstream codons facilitated read-through. This region was then compared with other viruses known to use suppression for expression of their *pol* gene. The comparison revealed that there was a great deal of conservation in the sequence and secondary structure (pseudoknot) among the viruses that lent support to the region's importance in signaling read-through. Mutations that disrupted the formation of the stem of the predicted pseudoknot abolished read-through and compensatory mutations restored read-through ability, confirming that the pseudoknot played some role in the read-through signal. In addition, the octanucleotide spacer region between the leaky terminator and the beginning of the pseudoknot was examined for its role in read-through signaling. Mutational analysis of the region demonstrated that a purine rich spacer was also important for signaling read-through of the UAG of MuLV.

The only other unique viral read-through signal characterized thus far has been for the UAG of the coat protein of BYDV-PAV (34). The elucidation of this signal began with the observation that six to fifteen nucleotides downstream of the leaky stop in *Luteoviruses*, seven to sixteen copies of a C rich sequence (CCN-NNN) occurred. A series of frameshift mutations within the repeat region of BYDV revealed that when fourteen of the sixteen repeats were taken out of frame, read-through ceased in wheat germ extract, supporting the repeat's role in signaling. In addition to the repeat proximal to the stop codon, deletion of a region nearly seven hundred nucleotides downstream of the terminator also prevented read-through from occurring. The boundaries of this distal

sequence were identified by a series of 5' and 3' deletions, showing the signal to reside between nucleotides 697 and 758 relative to the terminator.

The two most proximal 3' codons relative to the terminator were also subjected to mutational analysis to determine their involvement in signaling read-through for BYDV. The analysis revealed that no mutation could be found that altered read-through, suggesting that this region has no part in the signal.

The proximal and distal signals were then tested in oat protoplasts using a glucuronidase expression vector. These constructs demonstrated that the distal signal still functioned at 50% efficiency when placed 1.7 kb downstream of the terminator, and that read-through was still signaled at approximately 9% of wild-type even when all but two of the proximal repeats were removed.

The above studies illustrate the unique viral read-through signals known thus far. However, how many different kinds of signals function in read-through is not known. The only attempt at a larger classification scheme for read-through signals was described in a recent review article by Beier and Grimm, where they separate many, but not all, read-through viruses into three types (64). Their tripartite classification scheme will be explored in more detail in Chapter IV.

Here, I describe a comparative analysis of 91 viral sequences proximal to read-through codons that revealed similarities among the sequences immediately following the stop codons. This comparison led to the identification of six distinct signals accounting for most of the sequences examined. These signals were then tested for their ability to signal read-through using the dual luciferase reporter system described above.

CHAPTER II: METHODS

Collection and Examination of Viral Sequences

Using the Taxonomy browser at the National Center for Biotechnology Information's (NCBI) web site (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/), several representative nucleotide sequences from each currently accepted International Committee on Taxonomy of Viruses (ICTV) genus were examined to determine whether any members of that genus contained a read-through stop codon. In addition, an Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) keyword search was performed for the terms "read-through" or "transl_except", and the resulting hits were further examined.

The uniqueness of each sequence was determined using criteria that allowed different strains of the same virus to be included, but excluded any sequences with identical names or known aliases as determined by their NCBI taxonomic entries. The 82 nt long segment of sequence from the 19th nucleotide 5' to the 60th nucleotide 3' of the leaky stop codon, was extracted from each identified distinct sequence. The segments were then compared to one another, and one sequence was excluded from any pair of sequences that shared >90% identity. Sequence controls were obtained by extracting the 82 nt region surrounding the in-frame, non-leaky stop codon downstream of each leaky stop codon.

The triplet immediately 3' of the stop codon (+1 triplet), was used to divide the sequences into six groups. The nonrandomness associated with each nucleotide position in each group was examined by chi-square (χ^2) analysis (Figure 1). The secondary structure 3' of the stop codon was examined using mFold (65,66). The folding was

1. Hypothetical raw sequence information with nucleotide position Y highlighted

Sequence #		Sequence #	
1	AAUCCGCU	9	GCGCUAUG
2	GGCAAUGC	10	AAAAUGCC
3	GAUAUAUA	11	CGUAAUGA
4	GCAUGCAU	12	GGUUUAGC
5	CCCCCCCG	13	GAAUCGAA
6	AUAUCGAA	14	GGGGUAUA
7	GUUUAUCG	15	GAAAAUUU
8	GAUGCCGU	16	GUCAUUGG

2. Contingency tables for hypothetical nucleotide position Y

	A	U	C	G	Total
# of sequences that have X nucleotide at Y position	3	0	2	11	16
# of sequences that do not have X nucleotide at Y position	13	16	14	5	48
Total	16	16	16	16	64

	A	U	C	G
# with X at Y				
Actual	3	0	2	11
Predicted (25% of Total)	4	4	4	4
# w/o X at Y				
Actual	13	16	14	5
Predicted (75% of Total)	12	12	12	12

3. Let fi represent the actual frequency in cell i and let fi^* represent the predicted frequency for cell i.

4. Then, if there are n cells, the test statistic would look like:

$$S = \sum_{i=1}^{n} \frac{(f_i - f_i^*)^2}{f_i^*}$$

5. Since there are eight cells, the value of the test statistic in this case is:

$$S = \frac{(3-4)^2}{4} + \frac{(0-4)^2}{4} + \frac{(2-4)^2}{4} + \frac{(11-4)^2}{4} + \frac{(13-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(14-12)^2}{12} + \frac{(5-12)^2}{12}$$

$$= 23$$

5. This procedure was repeated for each nucleotide position in each group.

Figure 1: Chi-square method of statistical analysis as applied to the sequence groups

simulated at both 37°C and 25°C, and limitations were set that allowed G:U pairing, but did not allow any binding to occur with the stop codon or the sequence 5' of it.

Transfection and Assessment of the Reporter Vector p2luc Constructs

All constructs made from the dual luciferase reporter vector p2luc (Accession # AF043450) (61) were constructed by Xuifen Li from Dr. John Atkins Laboratory at the University of Utah and given to me for testing. For analysis, each of the seventeen 18 bp sequences, representing groups 1, 2, 3 and 4 of the identified read-through groups, were inserted between the *Bam*HI and *Sal*I restriction sites. The insert of each construct was synthesized as a pair of complementary oligonucleotides, such that the coding strand sequence read 5'GATCC-CCC-AAA-WWW-XXX-XXX-CAG$^{3'}$, and the non-coding strand sequence read $^{5'}$TCGAC-CTG-YYY-YYY-ZZZ-TTT-GGG$^{3'}$. GATCC and TCGAC were the complementary sticky ends of the *Bam*HI and *Sal*I sites, respectively. The WWW was either TAG or TGA for the test sequences or CAG or TGG for the control sequences and the Z's represent the complement of the corresponding W. The X's represent the +1 and +2 codon nucleotide positions and the Y's represent the complement of the corresponding X. Each pair of oligonucleotides were then annealed and separately ligated into the digested p2luc vector. (Figure 2).

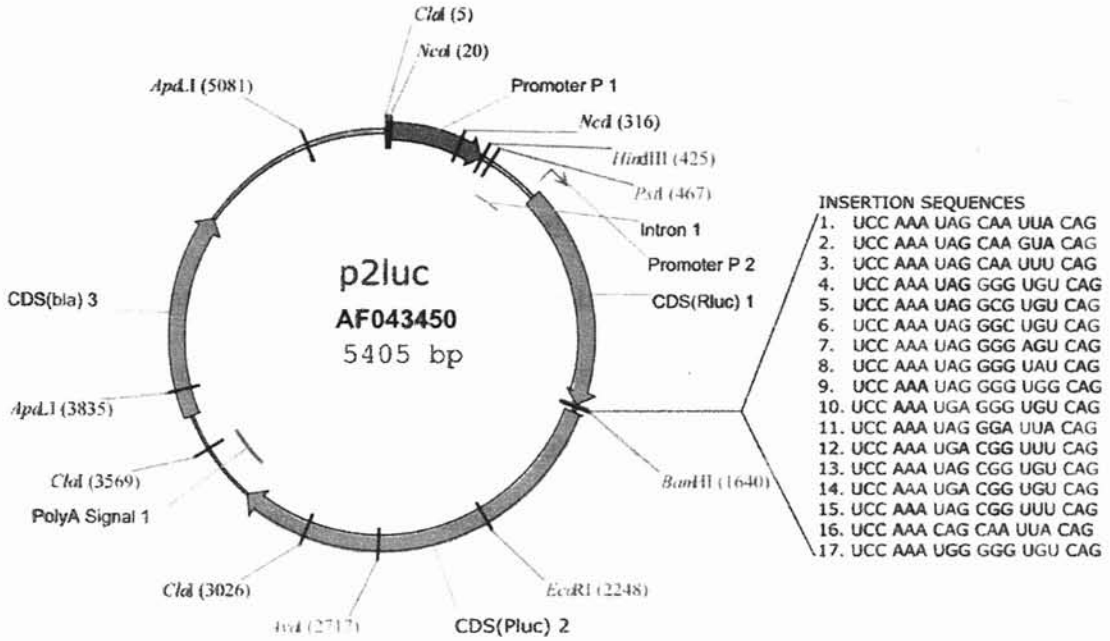

Figure 2: Diagram of p2luc constructs (modified from Vector NTI)

Optimization of Luciferase Assay Conditions

To determine the optimum conditions for the luciferase assay, two different mammalian cell lines, human embryo kidney number 293 (Hek 293) and simian fibroblast CV-1 (COS), were each tested with two different transfection reagents, Lipofectin™ and Lipofectamine™ (Gibco BRL). Each cell line was taken from a stock maintained in a 4x6x1'' culture flask at 37°C in 10 ml Dulbecco's Modified Eagle Medium (Gibco BRL) culture medium containing 10 % fetal bovine serum (DMEM-10%-FBS), and was prepared for transfection under a laminar flow hood by the following procedure: The DMEM-10%-FBS was removed using a Pasteur pipette under vacuum. The cells were then washed twice by adding 10 ml of phosphate buffered saline and removing it using a Pasteur pipette under vacuum. The cells were then incubated at 37°C for 10 min with 4 ml of Trypsin-EDTA (Gibco BRL) solution to dissociate them from the bottom of the culture flask. Following the incubation, 10 ml of DMEM-10%-FBS was added to the flask and the whole suspension was transferred to a 50 ml capped centrifuge tube and centrifuged for 2 min at 200 xg. All except 500 µl of the supernatant was removed, and the pellet was resuspended and then mixed with 10 ml of DMEM-10%-FBS. The concentration of cells was then determined using a hemocytometer. The cells were diluted to $1x10^5$ cells/ml before splitting them into 200 µl aliquots in a 48 well micro titer plate. The micro titer plate was then placed in an incubator at 37°C (24 hours for COS and 48 hours for Hek 293) to allow the cells to attach to the bottom of the plate and divide.

The p2luc constructs were prepared for transfection by first mixing 20 µl of serum-free DMEM (DMEM-SF) with 0.6 µl of Lipofectin™ or Lipofectamine™ per

sample. This mixture was allowed to incubate at room temperature for 45 min before mixing it with 20 µl of DMEM-SF containing 200 ng of construct DNA. Fifteen minutes after the mixing, 160 µl of DMEM-SF was added for a total volume of 200 µl per sample.

After the cell incubation was complete, the medium in each well was replaced with 200 µl of the transfection solution, and the plate was allowed to incubate overnight at 37°C. The medium was then replaced with 200 µl of DMEM-10%-FBS and the incubation was continued another 24 hours. After 24 hours the medium was replaced with 100 µl of 1X Lysis Buffer from the Dual Luciferase Assay System (Promega) and the plate was placed at -80°C until frozen. The plate was then thawed and 70 µl of each sample was transferred to a 96 well microtiter plate in preparation for assaying.

The assays were performed using a Dynax 96 Luminometer using a 2 sec delay between the injection of the reagent, either luciferase assay reagent (LAR II) for *Photinus* luciferase or Stop & Glow® for *Renilla* luciferase (both from Dual Luciferase Assay System), and the beginning of each reading with a reading duration of 10 sec. The raw results in relative light units (RLU) were then converted into average percent read-through (Ave % RT) (61). The COS cells transfected with Lipofectamine™ showed on average 100 fold higher expression levels than any other combination and displayed the lowest experimental error, so they were used for the repeat experiments.

CHAPTER III: RESULTS

Collection and Examination of Viral Sequences

Of the viral genera examined, 23 had read-through stop codons reported in their genome annotations. These 23 genera yielded 157 individual sequences for screening. The uniqueness of each sequence was determined as described in Materials and Methods, eliminating 66 sequences and leaving 91 unique sequences to be analyzed.

Since the nucleotides immediately 3' of the leaky stop codon were previously implicated in read-through (31,32,34,63,67-71), the sequences were first categorized according to these nucleotides. It was found that 6 of the 64 possible nucleotide triplets accounted for 90% of the triplets in the +1 position of the read-through sequences, but were present in only 2% of the non-read-through control sequences. In contrast, the six most frequent triplets found in the +1 position of the non-read-through control groups could categorize only 35.1% of the control sequences. Each one of the six triplets, CAA, CGG, GGG, GGA, GUA and CUA, formed the basis of a sequence group into which all sequences containing that +1 triplet were placed (Table 1).

Sequences in the +2 triplet position also were distributed non-randomly with the five most frequent accounting for 70%. These triplets were distributed among all the +1 groups. However, for five of the six +1 groups, there was a +2 triplet that was mostly non-random for each +1 triplet sequence: CAA-UUA (88%); CGG-UUU (55%); GGG-UGC (53%); GGA-GGC (66%); and GUA-GAC (80%). The sixth group, CUA, did not show any common +2 triplet sequence.

Group 1 (CAA)									
Genus	**Virus**	**-3**	**-2**	**-1**	**Stop**	**+1**	**+2**	**+3**	**Accession #**
Tobamo	*Turnip vein clearing*	GGG	GUC	CAA	UAG	CAA	UUA	AUU	U03387
	Chinese rape mosaic	GGU	ACC	CAA	UAG	CAA	UUA	CAG	U30944
	Tobacco mosiac	GGU	ACU	CAA	UAG	CAA	UUA	CAG	AF155507
	Tobacco mosaic (B935A)	GGA	ACA	CAA	UAG	CAA	UUA	CAG	AJ011933
	Tobacco mosiac (crucifer, Cg)	GGG	ACC	CAA	UAG	CAA	UUA	CAG	D38444
	Tobacco mosaic (Korean)	GGA	ACA	CAA	UAG	CAA	UUA	CAG	X68110
	Tobacco mosaic (K2)	GGU	ACU	CAA	UAG	CAA	UUA	CAG	Z92909
	Tobacco mosaic (OM)	GGA	ACA	CAA	UAG	CAA	UUA	CAG	D78608
	Tobacco mosaic (Tomato/L)	GGU	ACU	CAA	UAG	CAA	UUA	CAG	X02144
	Tobacco mosaic (crucifer, Russian)	GGG	AUC	CAA	UAG	CAA	UUA	CAG	Z29370
	Tobacco mild green mosaic (U2)	GGU	AGU	AGA	UAG	CAA	UUA	CAG	M34077
	Tobamovirus Ob	GUG	AGU	GCA	UAG	CAA	UUA	CAG	D13438
	Pepper mild mottle (S)	UCG	ACU	CAA	UAG	CAA	UUA	CAG	M81413
	Cucumber green mottle mosaic (SH)	CCU	ACC	AAA	UAG	CAA	UUA	AUG	D12505
	Cucumber green mottle mosaic (YODO)	UCC	CCC	AAA	UAG	CAA	UUA	AUG	AB015145
	Sunn-hemp mosaic	ACC	CAA	AAA	UAG	CAA	UUA	CAG	U47034
	Odontoglossum ringspot (Singapore)	GGG	AUC	UUA	UAG	CAA	UUA	CAG	U34586
	Cucumber fruit mottle mosaic	GGG	ACC	AAA	UAG	CAA	UUA	CAG	AF321057
Beny	*Beet necrotic yellow vein (S)*	CCC	GGA	CAA	UAG	CAA	UUA	GCU	D84411
	Beet soil-borne mosaic	CGC	ACC	AAU	UAG	CAA	UUA	AAU	AF061869
Pomo	*Beet soil-borne (Ahlum)*	UGG	GUU	GAA	UAG	CAA	UCA	ACU	U64512
	Broad bean necrosis	CCG	ACA	GCA	UAA	CAA	UUA	ACG	D86637
	Potato mop top (U)	GCU	GGU	GCA	UAG	CAA	UUA	ACC	D19613
	Beet virus Q-RNA2	ACC	GGC	UCA	UAG	CAA	UCA	AUU	AJ223597
UNCLASS	*Botrytis virus F*	GCU	GAA	CCA	UGA	CAA	UCA	CAG	AF238884
Group 2 (CGG)									
Furo	*Chinese wheat mosaic (Yantai)-RNA1*	UUC	GAC	AAA	UGA	CGG	UUU	GGG	AJ012005
	Chinese wheat mosaic (Yantai)-RNA2	AGG	UUC	GAG	UGA	CGG	GAU	GGC	AJ012006
	European wheat mosaic	UUG	GCG	AAA	UGA	CGG	UUU	GGG	AJ132576
	Oat golden stripe-RNA1	AAU	CAG	AAA	UGA	CGG	UUU	GGG	AJ132578
	Oat golden stripe-RNA2	GGT	AGU	GCC	UGA	CGG	GGC	GGC	AJ132579
	Soil-borne rye mosaic (O)	UUG	GUG	AAA	UGA	CGG	UUU	GGG	AF146280
	Soil-borne wheat mosaic (Japanese)	AAC	GGG	AAA	UGA	CGG	UUU	GGG	AB033689
	Soil-borne wheat mosaic (US-N)-RNA1	CUU	ACU	AAA	UGA	CGG	UUU	GGG	L07937
	Soil-borne wheat mosaic (US-N)-RNA2	GGU	UCG	AGU	UGA	CGG	GAC	GGC	L07938
	Soil-borne wheat mosaic (UK-Kent)	GGU	UCG	AGU	UGA	CGG	GAC	GGC	AJ298070
	Soil-borne wheat mosaic (UK-Wiltshire)	GGU	ACG	AGU	UGA	CGG	GAC	GGC	AJ298069
	Sorghum chlorotic spot	CAU	ACC	AAA	UGA	CGG	UUU	GGG	AB033691
Pomo	*Beet virus Q-RNA1*	UCU	GUU	CAA	UAA	CGG	UUU	GGG	AJ223596
Peculo	*Peanut clump*	CAG	ACC	AAA	UGA	CGG	UUU	GGG	X78602
Tobra	*Pepper ringspot (CAM)*	GCU	GCC	UUA	UGA	CGG	UGU	CGG	L23972
	Tobacco rattle (North American)	ACC	GUC	UUA	UGA	CGG	UUU	CGG	AF034622
	Pepper ringspot (CAM)	GCU	GCC	UUA	UGA	CGG	UGU	CGG	L23972
	Pea early browning	GCU	AUG	AAA	UGA	CGG	UGU	CGG	X14006
Colti	*Colorado tick fever*	GGC	UGC	UGU	UGA	CGG	UGU	UGG	AF000720
Alpha	*Venezuelan equine encephalitis (83U434)*	CAA	CAA	CAA	UGA	CGG	UUU	GAC	U55362
	Venezuelan equine encephalitis (68U201)	CAA	CAG	CAA	UGA	CGG	UUC	GAC	U34999

Genus	Virus	-3	-2	-1	Stop	+1	+2	+3	Accession #
Group 3 (GGG)									
Aureus	*Pothos latent (pigeonpea)*	GAU	GUC	UAC	UAG	GGG	UGC	CUA	AJ243370
Machlomo	*Maize chlorotic mottle*	GAG	UUG	AAA	UAG	GGG	UGU	UCU	X14736
Enamo	*Pea enation mosaic (At+)*	GCC	UCC	CUC	UGA	GGG	GAC	GAC	Y09099
Carmo	*Cardamine chlorotic fleck*	UUU	GUC	CGC	UAG	GGG	UGC	UUA	L16015
	Carnation mottle (Shanghai)	UUU	CCC	AAA	UAG	GGG	GGC	CUG	AF192772
	Galinsoga mosaic carmovirts	CUG	GGC	AAA	UAG	GGG	UGC	CUU	Y13463
	Hibiscus chlorotic ringspot	CCC	GUG	AAA	UAG	GGG	UGC	CUU	X86448
	Japanese iris necrotic ring	UUC	UCC	AAC	UAG	GGG	UGC	CUC	D86123
	Melon necrotic spot	UUG	GUC	AAC	UAG	GGG	UGC	CUG	M29671
	Saquaro cactus	UAC	CAC	AAA	UAG	GGG	UGC	CUA	U72332
	Turnip crinkle	UUU	GUC	CGC	UAG	GGG	UGC	UUG	M22445
Necro	*Leak white stripe*	CAU	GCC	AAA	UAG	GGG	GGC	CUA	X94560
	Tobacco necrosis (Type A)	CGG	UCC	AAA	UAG	GGG	UGC	CCU	X58455
Panico	*Panicum mosaic*	UUU	GCC	AAG	UAG	GGG	UGU	AUC	U55002
Retro C	*Baboon endogenous (M7)*	GAC	AGC	GAA	UAG	GGG	GGU	CAG	D10032
	Murine leukemia (SL3-3)	UUA	GAC	GAC	UAG	GGG	UGU	CAG	AF169256
UNCLASS	*Carrot red leaf luteovirus assoc RNA*	UAC	CGU	AAA	UAG	GGG	GGC	CUU	AF020616
Group 4 (GGA)									
Tombus	*Tomato bushy stunt (statice)*	GGU	GUC	AAA	UAG	GGA	GGC	CUA	AJ249740
Necro	*Tobacco necrosis (D)*	UGG	GAG	AAA	UAG	GGA	GGC	CUA	U62546
Lenti	*Simian immunodeficiency*	ACA	GAU	AAA	UAG	GGA	UUA	CUA	M92675
Group 5 (GUA)									
Luteo	*Barley yellow dwarf (PAV-I11)*	ACG	GCC	AAA	UAG	GUA	GAC	UCC	AF235167
	Barley yellow dwarf (PAV-129)	ACG	GCC	AAA	UAG	GUA	GAC	UCC	AF218798
	Barley yellow dwarf (SGV-U)	AAC	CCC	AAA	UAG	GUA	GAC	CCC	U06866
	Barley yellow dwarf (MAV)	ACU	CCC	AAA	UAG	GUA	GGC	UCC	D11028
	Barley yellow dwarf (SDV)	AAU	GCU	AAA	UAG	GUA	GAC	GGA	L24049
	Barley yellow dwarf (RPV)	AAC	CCA	AAA	UAG	GUA	GAC	GCG	L25299
	Beet western yellow (FL1)	AAC	CCC	AAA	UAG	GUA	GAC	GAG	X13063
	Sugarcane yellow leaf	AAU	CCC	AAA	UAG	GUA	GGC	GAC	AF157029
Polero	*Potato leaf roll*	AAC	CCC	AAA	UAG	GUA	GAC	UCC	X14600
Clostero	*Cucurbit aphid-borne yellows*	AAC	CCG	AAA	UAG	GUA	GAC	GGC	X76931
Group 6 (CUA)									
Alpha	*Sindbis (XJ-160)*	ACU	GAA	UAC	UGA	CUA	ACC	GGG	AF103728
	Sagiyama	AAC	CAG	UCC	UGA	CUA	GGC	AGG	AB032553
	Sindbis-like	ACC	GAA	UAC	UGA	CUA	ACC	GGG	AF103734
	Onyong-nyong	GAA	GAG	UUA	UGA	CUA	GAC	AGA	AF079456
	Middleburg	ACG	UCA	GCA	UGA	CUA	GAC	CGG	J02246
Cricket P-L	*Plautia stali intestine*	GAA	GAA	AGC	UGA	CUA	UGU	GAU	AB006531
Ungrouped Sequences									
Retro C	*Feline leukemia**	UUA	GGA	GAU	UAG	GAG	AGU	CAG	AF052723
	*Spleen necrosis**	GAA	UUA	CAA	UGA	GGC	CGU	CAG	M54993
Allolevi	*Bacteriophage M11**	CCG	GCG	UAU	UGA	ACU	GCG	CUU	AF052431
	*Bacteriophage Q-beta**	CCA	GCG	UAU	UGA	ACA	CUG	CUC	M99039
	*Bacteriophage SP**	CCA	GCC	UAC	UGA	GCG	GCG	UUA	X07489
SERUM GROUP IV OF LEVIVIRIDAE	*Bacteriophage NL95*	CCU	GCA	UAC	UGA	GCA	GCG	UUA	AF059243
	*Bacteriophage MX1**	CCG	GCG	UAC	UGA	ACU	GCA	CUU	AF059242
Lenti	*Feline immunodeficiency*	UUA	AGA	AAU	UGA	UAU	UAA	CCA	D83733
Furo	*Chinese wheat mosaic (Rongcheng)*	CAC	UAU	GAA	UGA	GUG	UUA	CAC	AJ271839
Machlomo	*Maize chlorotic mottle*	AAU	UUC	AAC	UGA	GCU	UUA	GUG	X14736

Table 1: Categorization of viral read-through sequences based on the +1 triplet. Viruses in the Ungrouped Sequences category are those which are indicated in the text as containing at least part of the Luteovirus read-through signal.

Examination of sequence variability within the groups using chi-square analysis (Figure 3) showed that both the +1 and, to a lesser extent, +2 triplets in the read-through groups were much more non-random than their counterparts in the corresponding control groups. Read-through groups 4, 5 and 6 contain other nucleotides that have chi-square values as high as, or nearly as high as, their respective nucleotides in the +1 triplet. For group 4, the high level of background in both the read-through and control groups can be accounted for by the fact that this group only contains three viral sequences. For group 5, which consists mostly of *Luteoviruses*, the other highly non-random nucleotides are expected because of the CCN-NNN repeat that has been shown to be important in signaling read-through for *Luteoviruses* (46). This repeat is present in the other sequences in group 5, and in six of the nine ungrouped sequences, suggesting that this repeat may also play a role in signaling read-through in other genera. In group 6, the equality of other chi-square values with that of the +1 triplet is due to the higher level of sequence identity among the coding regions of the read-through proteins of the *Alpha viruses* as compared to members of other groups. This is evident by comparing the chi-square values 3' of the leaky terminator in the read-through group and the chi-square values 5' of the read-through protein stop codon in the control group with those 3' of the control group.

The terminal dipeptide has been implicated in translation termination efficiency (27,68,72). The chemical characteristics of the penultimate amino acid have been shown to influence the efficiency of termination, with basic residues yielding more efficient read-through in *Saccharomyces cerevisiae* (68), while acidic and hydrophobic

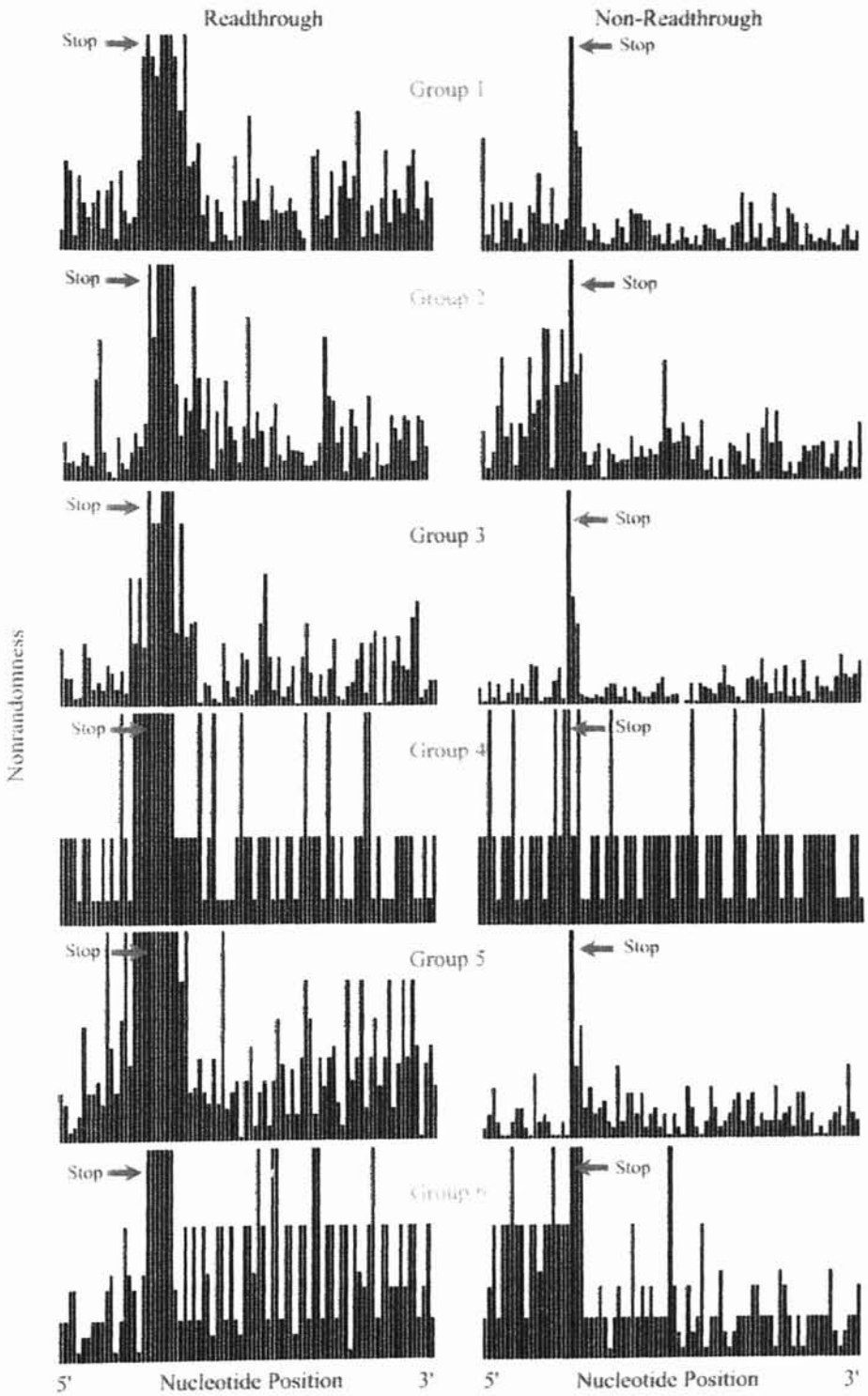

Figure 3: Examination of non-randomness in read-through groups: Each data set represents the 82 nucleotide region surrounding either the leaky (left column) or non-leaky (right column) stop codon for each of the six +1 triplet groups. For each data set, the X-axis represents the individual nucleotide position from the -19^{th} to the $+60^{th}$ relative to the stop codon, and the Y-axis is a measure of the non-randomness associated with each nucleotide position in the form of a Chi-square value.

residues give higher read-through in *E. coli* (72). Greater read-through efficiency is also associated with a higher likelihood of the ultimate amino acid participating in the formation of α-helices or β-sheets in *E. coli* (27). I compared the properties of the terminal dipeptides in both my sample and control groups and found no bias in the chemical characteristics of the penultimate amino acid, or in the α-helical or β-sheet propensities of the ultimate amino acid, suggesting that these 5' signals are not utilized to facilitate viral read-through. However, comparison of the 5' sequences revealed that there was a preference for adenine in the penultimate and ultimate nucleotide positions, accounting for 76% and 71% respectively, of the total nucleotides in those positions compared to only 28% and 25%, respectively, in the non-readthrough control groups.

The secondary structure 3' of the stop codon was examined for stem-loop or pseudoknot structures reported to be important for efficient read-through in some viruses (63). Although some of the sequences examined displayed feasible structures, no consistent significant similarities were found with either the reported structures or with each other.

Construct Analysis using Luciferase Reporter System

Although all of the different cell line/transfection reagent combinations support the same conclusions—they display the same overall patterns—the results discussed in the text are derived from the COS/Lipofectamine™ assays for the reasons described in the last paragraph of Chapter II. However, the results of all the cell/transfection reagent combinations are displayed in Tables 2-6. The construct inserts were designed either to be exact replicas of the +1 and +2 triplets of the sequence groups, as was the case for constructs 1, 4, 11, 12 and 14 (groups 1, 3, 4, 2 and 2 respectively), or derivatives of

those sequences with single nucleotide substitutions, as was the case for the remaining ten constructs. Each of the fifteen constructs inserted into the p2luc vector was tested for its ability to facilitate read-through expression of *Photinus* luciferase. The raw output of the COS/Lipofectamine™ assays was converted to Ave. % RT, which is displayed in Figure 4. The Ave. % RT of the other combinations are shown in Figure 5.

The sequence for construct 1 was taken directly from group 1 (CAA-UUA) and constructs 2 and 3 each make a single substitution in that sequence (CAA-GUA and CAA-UUU, respectively). While the U to G substitution at position +4 had little effect on read-through, the A to U substitution at position +6 significantly reduced the level of read-through observed. These results are consistent with those obtained by Skuzeski et al. examining the TMV read-through context in vivo (32) and the in vitro work of Zerfass and Beier (33).

Constructs 4-10 are all based on group 3 (GGG). None of the single nucleotide substitutions made in the +1 and +2 triplets in constructs 5-8 had any significant effect on read-through levels, however in construct 9, where the +6 nucleotide U was substituted with a G, read-through increased significantly. In construct 10, where the UAG stop codon was changed to a UGA, the read-through rose approximately 1.7 fold. A difference in read-through between identical constructs with different stop codons also occurred in constructs 12-14, which are based on group 2 (CGG); both the CGG-UUU and CGG-UGU constructs showed significantly higher levels of read-through when placed downstream of a UGA terminator compared to their UAG counterparts. Groups 5 and 6 were not tested because of limitations on the number of constructs that could be made.

When comparing the constructs that represent the non-substituted read-through group sequences, there also appeared to be variation in the percentage of read-through each sequence group facilitated. Group 1 showed the highest levels (3.2% +/- 0.4), followed by both group 2 constructs, themselves displaying variation (2.3% +/- 0.3 and 1.9% +/- 0.2). Group 4 had the third highest read-through levels (1.1% +/- 0.1), and group 3 had the lowest of the groups tested (0.8% +/- 0.05).

Table 2: Raw and converted results from the p2luc construct assays for COS/Lipofectamine™ (Set 1)
fluc: firefly (*Photinus*) luciferase; rluc: *Renilla* luciferase; RLU: relative light units; C: control

Cell Type	COS Set 1						
Reagent	Lipofectamine						
Construct #	**fluc: RLU**	**rluc: RLU**	**fluc/C: RLU**	**rluc/C: RLU**	**%RT**	**Ave %RT**	**St Dev**
1	3388.4	201470	245271	536914	3.68	3.228002	0.4261
1	2069.0	159691	245271	536914	2.84		
1	2791.5	193001	245271	536914	3.17		
2	3262.5	253052	245271	536914	2.82	3.246435	0.5408
2	3649.2	260915	245271	536914	3.06		
2	5102.7	289727	245271	536914	3.86		
3	781.0	192975	245271	536914	0.89	0.844007	0.0363
3	725.4	192501	245271	536914	0.82		
3	651.3	173600	245271	536914	0.82		
4	526.5	166656	245271	536914	0.69	0.73447	0.0445
4	957.1	268447	245271	536914	0.78		
4	997.2	298456	245271	536914	0.73		
5	283.7	94802	245271	536914	0.66	0.726105	0.0936
5	480.7	126462	245271	536914	0.83		
5	488.8	154832	245271	536914	0.69		
6	438.4	134491	245271	536914	0.71	0.721627	0.0121
6	768.2	228640	245271	536914	0.74		
6	380.6	116385	245271	536914	0.72		
7	870.9	233992	245271	536914	0.81	0.659321	0.1378
7	262.9	94141	245271	536914	0.61		
7	377.7	149812	245271	536914	0.55		
8	462.1	159887	245271	536914	0.63	0.74003	0.0983
8	470.7	124845	245271	536914	0.83		
8	370.1	106312	245271	536914	0.76		
9	663.0	161013	245271	536914	0.90	1.022538	0.1052
9	732.3	149067	245271	536914	1.08		
9	834.4	167452	245271	536914	1.09		
10	1308.7	293473	146878	395465	1.20	1.233213	0.1453
10	1225.8	237093	146878	395465	1.39		
10	785.0	190939	146878	395465	1.11		
11	1021.6	195882	245271	536914	1.14	1.054888	0.0752
11	804.2	173932	245271	536914	1.01		
11	720.9	156129	245271	536914	1.01		
12	2018.3	263923	146878	395465	2.06	1.878549	0.1579
12	1581.9	241206	146878	395465	1.77		
12	2453.5	364808	146878	395465	1.81		
13	2412.8	447942	245271	536914	1.18	1.102824	0.0801
13	1167.0	250603	245271	536914	1.02		
13	1284.9	253412	245271	536914	1.11		
14	2943.8	327995	146878	395465	2.42	2.318777	0.2684
14	2802.1	374376	146878	395465	2.02		
14	3807.1	406030	146878	395465	2.52		
15	2021.5	432977	245271	536914	1.02	1.058233	0.0318
15	1127.3	228159	245271	536914	1.08		
15	1362.9	278537	245271	536914	1.07		

Table 3: Raw and converted results from the p2luc construct assays for COS/Lipofectamine™ (Set 2) fluc: firefly (*Photinus*) luciferase; rluc: *Renilla* luciferase; RLU: relative light units; C: control

Cell Type	COS Set 2						
Reagent	Lipofectamine						
Construct #	**fluc: RLU**	**rluc: RLU**	**fluc/C: RLU**	**rluc/C: RLU**	**%RT**	**Ave %RT**	**St Dev**
1	4086.6	140637	182283	226275	3.61	2.848	0.660
1	1582.7	81341	182283	226275	2.42		
1	1365.0	67228	182283	226275	2.52		
2	3171.7	172751	182283	226275	2.28	2.757	0.686
2	4653.9	163023	182283	226275	3.54		
2	2977.5	150915	182283	226275	2.45		
3	819.4	119181	182283	226275	0.85	0.858	0.011
3	798.3	113817	182283	226275	0.87		
3	867.9	126727	182283	226275	0.85		
4	438.6	73519	182283	226275	0.74	0.719	0.054
4	1056.9	172960	182283	226275	0.76		
4	1115.4	210733	182283	226275	0.66		
5	243.2	49247	182283	226275	0.61	0.589	0.022
5	228.8	49920	182283	226275	0.57		
5	269.1	56997	182283	226275	0.59		
6	389.2	84651	182283	226275	0.57	0.545	0.033
6	285.2	63544	182283	226275	0.56		
6	566.8	138329	182283	226275	0.51		
7	289.3	68856	182283	226275	0.52	0.508	0.012
7	645.6	160034	182283	226275	0.50		
7	792.7	196049	182283	226275	0.50		
8	308.1	74761	182283	226275	0.51	0.488	0.023
8	343.9	87455	182283	226275	0.49		
8	459.0	122482	182283	226275	0.47		
9	422.7	104530	182283	226275	0.50	0.500	0.005
9	580.4	145975	182283	226275	0.49		
9	482.0	118798	182283	226275	0.50		
10	523.2	132758	91495	145322	0.63	0.621	0.009
10	474.7	123431	91495	145322	0.61		
10	552.0	140179	91495	145322	0.63		
11	190.9	48414	182283	226275	0.49	0.486	0.019
11	473.1	116719	182283	226275	0.50		
11	251.1	67061	182283	226275	0.46		
12	1167.1	105233	91495	145322	1.76	1.737	0.025
12	1873.2	171235	91495	145322	1.74		
12	3925.7	364447	91495	145322	1.71		
13	1484.0	334235	182283	226275	0.55	0.523	0.039
13	727.8	167462	182283	226275	0.54		
13	406.7	105341	182283	226275	0.48		
14	1597.6	147698	91495	145322	1.72	1.565	0.133
14	1747.9	186648	91495	145322	1.49		
14	1598.1	170475	91495	145322	1.49		
15	541.7	124673	182283	226275	0.54	0.538	0.022
15	811.0	180206	182283	226275	0.56		
15	723.2	174295	182283	226275	0.52		

Table 4: Raw and converted results from the p2luc construct assays for COS/Lipofectin™
fluc: firefly (*Photinus*) luciferase; rluc: *Renilla* luciferase; RLU: relative light units; C: control

Cell Type	COS						
Reagent	Lipofectin						
Construct #	**fluc: RLU**	**rluc: RLU**	**fluc/C: RLU**	**rluc/C: RLU**	**%RT**	**Ave %RT**	**St Dev**
1	339.3	17414	9726	22696	4.55	3.224	1.151
1	81.6	7107	9726	22696	2.68		
1	93.3	8900	9726	22696	2.45		
2	190.1	11269	9726	22696	3.94	3.429	0.574
2	84.2	7003	9726	22696	2.80		
2	46.7	3074	9726	22696	3.55		
3	30.3	6331	9726	22696	1.12	1.120	0.030
3	33.6	7192	9726	22696	1.09		
3	29.4	5953	9726	22696	1.15		
4	21.6	5880	9726	22696	0.86	0.874	0.013
4	13.9	3694	9726	22696	0.88		
4	15.5	4096	9726	22696	0.88		
5	9.0	2326	9726	22696	0.90	0.857	0.050
5	4.8	1385	9726	22696	0.80		
5	10.4	2815	9726	22696	0.86		
6	17.4	4610	9726	22696	0.88	1.268	0.650
6	10.3	2662	9726	22696	0.91		
6	63.1	7299	9726	22696	2.02		
7	17.3	5017	9726	22696	0.81	0.866	0.052
7	17.6	4561	9726	22696	0.90		
7	30.3	7950	9726	22696	0.89		
8	20.9	4229	9726	22696	1.15	0.952	0.174
8	12.3	3255	9726	22696	0.88		
8	10.0	2837	9726	22696	0.82		
9	28.7	7553	9726	22696	0.89	0.905	0.020
9	34.6	8969	9726	22696	0.90		
9	22.4	5642	9726	22696	0.93		
10	32.7	7786	4674	12539	1.13	1.085	0.043
10	27.5	6788	4674	12539	1.09		
10	23.8	6148	4674	12539	1.04		
11	17.8	4879	9726	22696	0.85	0.854	0.029
11	29.4	7751	9726	22696	0.88		
11	31.7	8965	9726	22696	0.83		
12	22.2	2943	4674	12539	2.03	3.016	1.664
12	96.0	5215	4674	12539	4.94		
12	19.2	2476	4674	12539	2.08		
13	59.0	13890	9726	22696	0.99	0.954	0.051
13	39.4	9435	9726	22696	0.98		
13	37.6	9780	9726	22696	0.90		
14	94.0	11098	4674	12539	2.27	2.423	0.250
14	88.2	10363	4674	12539	2.28		
14	176.9	17498	4674	12539	2.71		
15	34.7	11838	9726	22696	0.68	0.853	0.151
15	35.6	9245	9726	22696	0.90		
15	48.5	11586	9726	22696	0.98		

Table 5: Raw and converted results from the p2luc construct assays for Hek293/Lipofectamine™
fluc: firefly (*Photinus*) luciferase; rluc: *Renilla* luciferase; RLU: relative light units; C: control

Cell Type	Hek 293						
Reagent	Lipofectamine						
Construct #	**fluc: RLU**	**rluc: RLU**	**fluc/C: RLU**	**rluc/C: RLU**	**%RT**	**Ave %RT**	**St Dev**
1	37.6	1400	1172	3099	7.10	7.044201	0.1546
1	41.9	1546	1172	3099	7.16		
1	43.2	1665	1172	3099	6.87		
2	44.7	1807	1172	3099	6.54	7.56034	0.89
2	67.5	2236	1172	3099	7.98		
2	83.8	2717	1172	3099	8.16		
3	7.5	1051	1172	3099	1.88	1.665189	0.1836
3	5.9	1006	1172	3099	1.55		
3	4.0	664	1172	3099	1.57		
4	4.3	945	1172	3099	1.21	1.212815	0.0056
4	6.6	1437	1172	3099	1.22		
4	5.4	1172	1172	3099	1.22		
5	4.2	904	1172	3099	1.24	1.234071	0.055
5	2.7	607	1172	3099	1.18		
5	3.9	805	1172	3099	1.28		
6	6.6	1436	1172	3099	1.21	1.20631	0.067
6	6.4	1483	1172	3099	1.14		
6	5.7	1192	1172	3099	1.27		
7	16.0	2496	1172	3099	1.70	1.226655	0.4126
7	9.7	2486	1172	3099	1.03		
7	6.9	1924	1172	3099	0.95		
8	5.4	1496	1172	3099	0.95	1.037794	0.073
8	7.5	1836	1172	3099	1.08		
8	6.1	1490	1172	3099	1.08		
9	5.7	1132	1172	3099	1.34	1.448225	0.126
9	5.8	1076	1172	3099	1.42		
9	5.8	968	1172	3099	1.59		
10	7.5	1419	653	2749	2.22	2.326617	0.0931
10	6.2	1101	653	2749	2.37		
10	9.7	1708	653	2749	2.39		
11	8.4	1491	1172	3099	1.49	1.546873	0.0968
11	8.2	1311	1172	3099	1.66		
11	7.2	1280	1172	3099	1.49		
12	22.6	2500	653	2749	3.80	3.41361	0.3483
12	20.1	2537	653	2749	3.32		
12	15.9	2142	653	2749	3.12		
13	11.8	1880	1172	3099	1.67	1.718766	0.1383
13	12.8	2095	1172	3099	1.61		
13	11.2	1574	1172	3099	1.88		
14	14.5	2363	653	2749	2.58	2.760701	0.2135
14	13.1	2044	653	2749	2.70		
14	17.6	2468	653	2749	3.00		
15	8.0	1299	1172	3099	1.63	1.586799	0.0429
15	8.9	1520	1172	3099	1.55		
15	12.3	2057	1172	3099	1.58		

Table 6: Raw and converted results from the p2luc construct assays for Hek293/Lipofectin™
fluc: firefly (*Photinus*) luciferase; rluc: *Renilla* luciferase; RLU: relative light units; C: control

Cell Type	Hek 293						
Reagent	Lipofectin						
Construct #	**fluc: RLU**	**rluc: RLU**	**fluc/C: RLU**	**rluc/C: RLU**	**%RT**	**Ave %RT**	**St Dev**
1	7.2	319	138	235	3.86	3.301	0.555
1	4.0	250	138	235	2.75		
1	4.4	227	138	235	3.29		
2	5.2	263	138	235	3.37	4.239	0.751
2	5.6	205	138	235	4.62		
2	15.9	571	138	235	4.73		
3	1.0	125	138	235	1.33	1.267	0.126
3	1.2	185	138	235	1.12		
3	1.2	153	138	235	1.34		
4	1.0	80	138	235	2.05	1.677	0.361
4	0.8	105	138	235	1.33		
4	1.0	104	138	235	1.65		
5	0.8	90	138	235	1.59	2.768	1.291
5	0.7	44	138	235	2.57		
5	0.6	26	138	235	4.15		
6	0.6	17	138	235	6.03	3.412	2.306
6	0.7	50	138	235	2.54		
6	0.8	78	138	235	1.67		
7	0.8	70	138	235	2.03	1.876	0.194
7	0.9	95	138	235	1.66		
7	0.7	64	138	235	1.94		
8	0.7	53	138	235	2.20	1.801	0.430
8	0.8	78	138	235	1.86		
8	1.0	122	138	235	1.35		
9	0.8	77	138	235	1.70	2.167	0.440
9	0.8	57	138	235	2.23		
9	0.7	46	138	235	2.57		
10	1.1	196	91	176	1.10	1.000	0.134
10	1.2	216	91	176	1.05		
10	1.7	387	91	176	0.85		
11	0.8	104	138	235	1.25	1.573	0.333
11	0.7	78	138	235	1.56		
11	0.7	58	138	235	1.91		
12	1.1	97	91	176	2.20	2.090	0.210
12	1.2	103	91	176	2.22		
12	1.2	129	91	176	1.85		
13	1.2	186	138	235	1.09	1.450	0.308
13	0.7	75	138	235	1.65		
13	0.9	94	138	235	1.60		
14	2.5	78	91	176	6.23	3.529	2.347
14	1.0	77	91	176	2.38		
14	1.4	138	91	176	1.98		
15	1.3	229	138	235	0.96	1.089	0.133
15	0.8	110	138	235	1.22		
15	1.2	187	138	235	1.09		

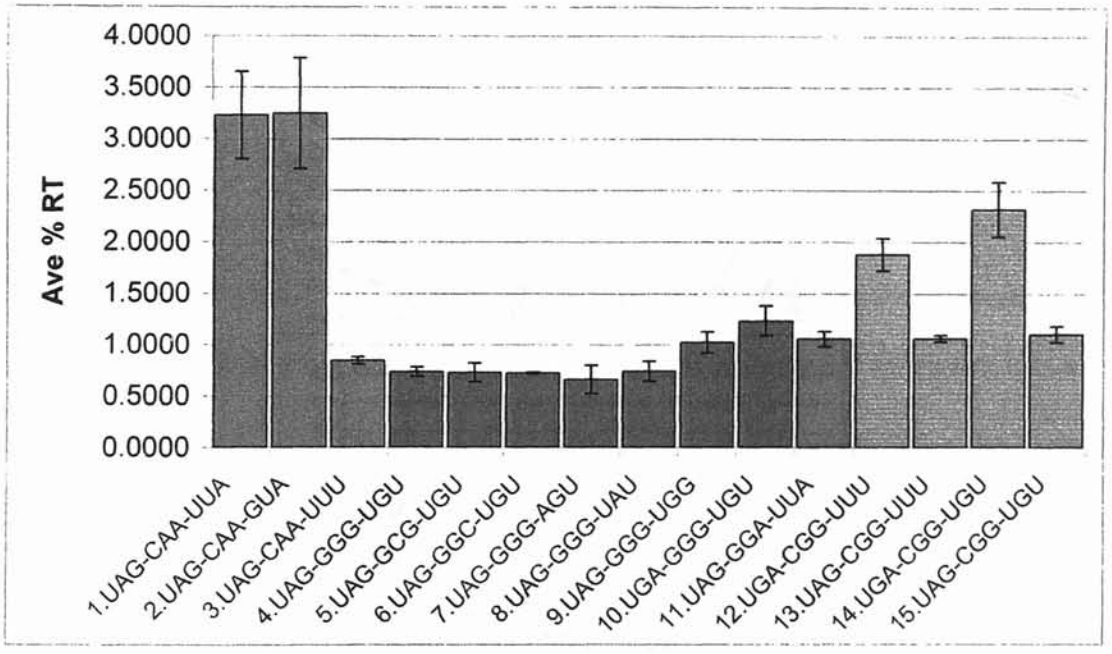

Figure 4: Read-through facilitated by p2luc constructs in COS cells using lipofectamine™: Each of the 15 sequences listed on the X-axis was tested for its ability to facilitate read-through of *Photinus* luciferase using the fusion vector p2luc. Sequences 1, 4, 11, 12 and 14 represent exactly the +1 and +2 triplet sequences from groups 1, 3, 4, 2 and 2 respectively. The other 10 sequences were either substitutions of a single nucleotide position (sequences 2, 3 and 5-9) or a substitution of the stop codon (sequences 10, 13 and 15). Constructs represented by blue, green, red and pink bars are derived from groups 1, 3, 2 and 4 respectively.

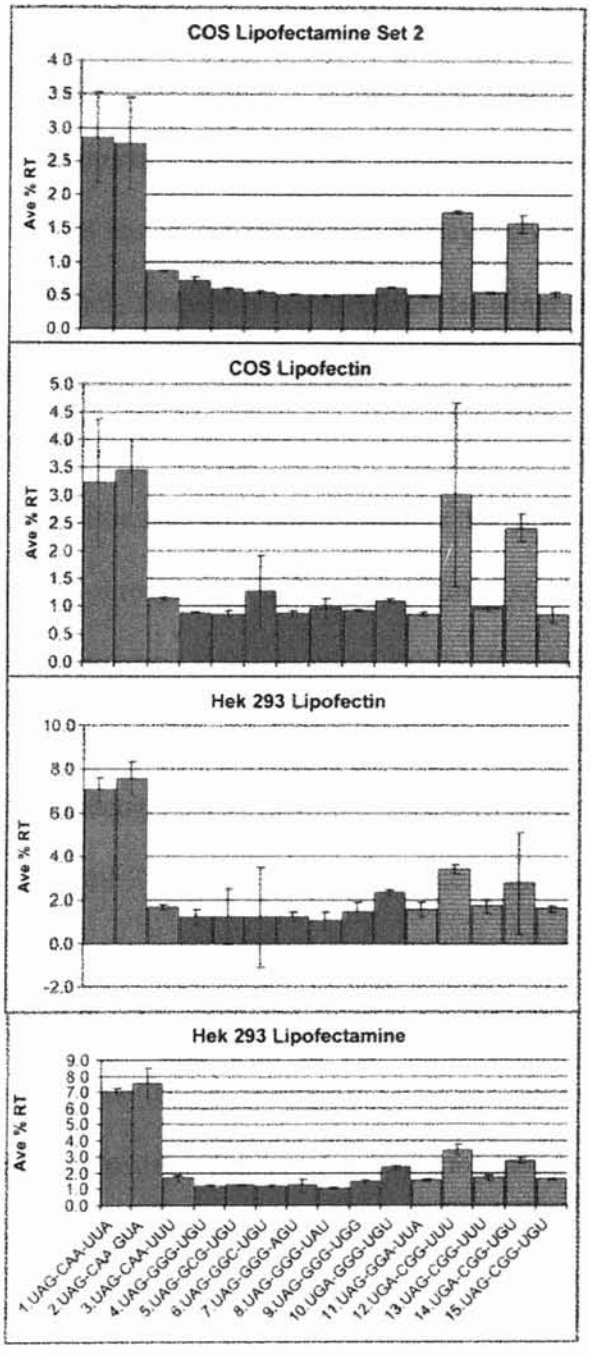

Figure 5: Read-through facilitated by p2luc constructs in all other cell transfection reagent combinations: Each of the 15 sequences listed on the X-axis was tested for its ability to facilitate read-through of *Photinus* luciferase using the fusion vector p2luc. Constructs represented by blue, green, red and pink bars are derived from groups 1, 3, 2 and 4 respectively.

CHAPTER IV: DISCUSSION

The degree to which the identity of the six nucleotides 3' of viral read-through stop codons is restricted in nature, is remarkable. The sequences examined are from RNA containing viruses in which mutation rates are notoriously high and different sequence combinations are undoubtedly frequently tested. Given the larger than triplet recognition in the release process (73), one would have expected the sequence 3' of "tight" (non-read-through) stop codons to be more restricted than leaky stop codons but this is not so, as the level of non-randomness associated with this position in the control groups is low (Figure 1). However, with the read-through stop codons, restriction extends to the sixth following nucleotide and is even pronounced at this position. While other recoding signals are known to be operative in some of the available sequences analyzed (31,32,34,74), and are likely involved, though unrecognized, in others, it is clear from the statistical and experimental analysis that the 3' hexanucleotide sequence is a major influence on read-through.

Stop codon recognition occurs in the ribosomal A-site. Stacking of the 3' adjacent base has an influence on codon interactions in the A-site and influences both termination and frameshifting (75-77). How the identity of up to six nucleotides affects read-through is less clear. In –1 frameshifting for yeast Ty3 there is provocative evidence that a local 3' effect, which extends to 13 bases, is due to mRNA pairing with rRNA in the pre A-site (78).

Read-through Signals

The number of nucleotides that appear to be necessary for the signaling of read-through vary. Group 1 adhered to the CAR-YYA formula found to be important for the in

vivo read-through of the UAG of TMV RNA (32) and the in vitro read-through of UAG, UAA and UGA in a TMV specific context (33). Group 1 is also in agreement with the CA(A/G)N(U/C/G)A consensus sequence found to facilitate read-through in *S. cerevisiae* (79) . The essential nucleotides in the spacer region between the stop codon and the beginning of the pseudoknot in MuLV (80) are mostly conserved in all of the members of Group 3. The *Luteovirus* proximal signaling sequence CCN-NNN is known to be necessary for read-through, whereas, the +1 and +2 triplet appear to have no importance (34). This *Luteovirus* signal appears, to various extents (two or more repeats), in every group 5 sequence, as well as appearing in six of the nine ungrouped sequences. Therefore, perhaps group 5 should be redefined using the CCN-NNN repeat as the criterion instead of the +1 triplet, leaving only three ungrouped sequences.

The type of stop codon appears to be a determinant for the +1 triplet groups, as almost all the groups are stop-codon specific. The only exceptions were *Broad bean necrosis virus* and *Botrytis virus F* in group 1, *Beet virus Q* in group 2 and *Pea enation mosaic virus* in group 3. Except for *Broad bean necrosis virus* in group 1 and *Beet virus Q* in group 2, the UAA stop codon does not appear in any of the sequences examined, which is consistent with UAA(A/G) being one of the most preferred termination sequences in eukaryotes (73). However, the context for UAG read-through in MuLV has been shown to work with UAA and UGA in vivo and in vitro (71). The TMV UAG read-through context also appears to function for both UAA and UGA in vitro (33). And the UGA read-through context of *Sindbis virus* can facilitate read-through for UAA and UAG as well (81). These data suggest that the stop codon dependence of the sequence

groups may be the result of some other factor and not a necessity for the functionality of the +1 triplet sequence.

The relative abundance of adenine in the penultimate and ultimate nucleotide positions relative to the leaky terminator suggests that the 5' context may also play a role in signaling read-through. However, previous studies report conflicting evidence as to the influence of adenine in these positions raising questions, at least at the nucleotide level, about the importance of these positions (80,82).

It is interesting to note that the individual groups are not host specific, since members of the *Furovirus*, *Coltivirus* and *Alphavirus* genera all appear in group 2, and infect plants, bacteria and mammals, respectively. Host non-specificity combined with the small number of groups that are sufficient to accommodate all but three of the examined sequences suggests one of two conclusions: that either there are only a limited number of sequences that can signal read-through, and the members of each group co-evolved the same sequence; or, less likely given the diversity of the members within a group, that the members of a group came from a common ancestor that possessed the signaling sequence or a precursor to it.

Although I was not able to categorize three of the examined sequences, this is likely a consequence of the limited number of complete viral genomes containing read-through stop codons that have been sequenced to date. Identification of more read-through sequences may lead to the addition of other groups of conserved sequences, allowing the ungrouped sequences here to be categorized, revealing the total number of signaling sequence groups and how they facilitate read-through of a stop codon.

A recent review article puts forth a different classification scheme than the one presented here (64). In this classification the sequences were divided into three Types. Type I represented plant viruses that contained the CAR-YYA consensus sequence of TMV (32), similar to my group 1. Type II contained both plant and animal viruses that had either a CGG or CUA +1 triplet 3' of a UGA terminator, where I have each of these triplets separated into groups 2 and 6 respectively. Type III is based on the linear, purine-rich octanucleotide found in the spacer region of MuLV, and while the spacer region is mostly conserved in many of the members of my group 3, in others it is not. This scheme, however, suffers from the major shortcoming of failing to examine, for unexplained reasons, all viral read-through sequences. This fact may explain why no Types can be found to compare with my group 4 or group 5.

Read-through Facilitated by the Selected Sequences in the p2luc

The varying amount of read-through observed with the different constructs suggests that some signals may play a larger role than others. The group 1 sequence CAA-UUA facilitates the highest amount of read-through of the tested sequences and approaches the level of ~5% reported by Skuzeski et al. for TMV (32). The constructs representing group 3, in contrast, show much lower levels than the ~5% level reported by Jamjoom et al. for MuLV (83). However, both the pseudoknot and the spacer region are known to play a role in read-through signaling in MuLV (63), a member of group 3. So perhaps additional signaling sequences also exist in the other members of group 3, and the GGG sequence is only part of a more complex signal. The presence of additional signals may also explain the presence of the CUA +1 sequence in group 5. As explained above, all the members of group 5 contain the CCN-NNN repeat known to be the

proximal read-through signal in *luteoviruses*. Perhaps the CUA sequence influences the efficiency of the repeat.

Inspection of the luciferase assay results also reveals some differences in the importance of the individual nucleotide positions in the constructs. In all the constructs assayed, and in all but four of the 91 sequences collected, either a C or G is in +1 nucleotide position. This supports evidence suggesting that the +1 position is critical for read-through in most systems, and that a C or G in that position is important for efficient read-through to occur (31,67,84). In fact, a C in the +1 nucleotide position is associated with all four constructs showing the highest levels of read-through.

The constructs that had substitutions in the +2 (construct 5), +3 (construct 6) and +4 (constructs 2 and 7) nucleotide position show no significant changes in the level of read-through facilitated, suggesting that these positions either have no role in signaling read-through or the nucleotides substituted are comparable to the ones that replaced them. Previous work with the *Tobacco rattle virus* (TRV) UGA context concluded that just a single nucleotide substitution was not sufficient to influence read-through (85). However, no substitutions to the +5 nucleotide were made in that study. My results show that the +5 U to G substitution in TRV's UGA context significantly increases read-through in my system. In contrast, my data also indicate that the same substitution at the +5 position had no influence on either the group 2 UAG construct series or on the group 3 construct series, suggesting that the importance of a position depends on the nature of the group and the stop codon. Alteration of the +6 position caused the most dramatic change in read-through of any subset tested. An A to U change in position +6 of group 1 decreased read-through 3.8 fold. In contrast, substituting a G for the +6 U in the group 3 construct

series increased read-through significantly. The terminator also appeared to have an influence on read-through levels in COS cells. For both the group 3 construct series and the group 2 construct series, higher levels of read-through were displayed when comparing constructs with identical +1 and +2 sequences, but having a UGA terminator instead of a UAG. These findings hint at the complexity of read-through signaling and demonstrate the need for additional constructs to further explore the importance of each nucleotide position.

Since the contexts tested here were taken from groups that include viruses whose hosts are in different kingdoms, read-through signaling mechanisms may be universal. Indeed, that the read-through analysis using luciferase reporter genes in mammalian COS cells is consistent with inferences derived primarily from viral sequences infecting plants supports this view. On the other hand, it may be possible that differences between the translation systems of the hosts would make the COS assay system used here a non-accurate representation of the performance of some of the contexts.

With the mechanism of translation termination only partially understood, it is difficult to determine with certainty what role the sequence groups described here play in altering that mechanism. The sequences in these groups could influence the binding equilibrium of either RF or aminoacyl-tRNA through direct contact or indirectly through interactions with the ribosome. It is unlikely that these groups have no role, because of their high level of non-randomness, their conspicuous under-representation in the control groups, and the ability of the plant viral sequences to facilitate read-through even using an animal based assay system. Surely as work continues to progress in translation termination, the role of these and other read-through signals will become apparent.

References

1. Crick, F.H. and Brenner, S. (1967) The absolute sign of certain phase-shift mutants in bacteriophage T4 *J Mol Biol*, **26,** 361-363.
2. Yanofsky, C. (1967) Gene structure and protein structure *Harvey Lect*, **61,** 145-168.
3. Khorana, H.G. (1977), *Nobel Lectures in Molecular Biology, 1933-1975*. Elsevier, pp. 303-331.
4. Jones, O.W., Jr. and Nirenberg, M.W. (1966) Degeneracy in the amino acid code *Biochim Biophys Acta*, **119,** 400-406.
5. Khorana, H.G., Buchi, H., Ghosh, H., Gupta, N., Jacob, T.M., Kossel, H., Morgan, R., Narang, S.A., Ohtsuka, E. and Wells, R.D. (1966) Polynucleotide synthesis and the genetic code *Cold Spring Harb Symp Quant Biol*, **31,** 39-49.
6. Holley, R.W. (1965) Structure of an alanine transfer ribonucleic acid *Jama*, **194,** 868-871.
7. Rich, A. and Kim, S.H. (1978) The three-dimensional structure of transfer RNA *Sci Am*, **238,** 52-62.
8. Moras, D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases *Trends Biochem Sci*, **17,** 159-164.
9. Gualerzi, C.O. and Pon, C.L. (1990) Initiation of mRNA translation in prokaryotes *Biochemistry*, **29,** 5881-5889.
10. McCarthy, J.E. and Brimacombe, R. (1994) Prokaryotic translation: the interactive pathway leading to initiation *Trends Genet*, **10,** 402-407.
11. Sergiev, P.V., Lavrik, I.N., Wlasoff, V.A., Dokudovskaya, S.S., Dontsova, O.A., Bogdanov, A.A. and Brimacombe, R. (1997) The path of mRNA through the bacterial ribosome: a site-directed crosslinking study using new photoreactive derivatives of guanosine and uridine *Rna*, **3,** 464-475.
12. Moazed, D. and Noller, H.F. (1990) Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16 S rRNA *J Mol Biol*, **211,** 135-145.
13. Moazed, D. and Noller, H.F. (1986) Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes *Cell*, **47,** 985-994.
14. Noller, H.F. and Chaires, J.B. (1972) Functional modification of 16S ribosomal RNA by kethoxal *Proc Natl Acad Sci U S A*, **69,** 3113-3118.
15. Dintzis, H. (1961) Assembly of the peptide chains of hemoglobin *Proc Natl Acad Sci U S A*, **47,** 247-261.
16. Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes *Gene*, **234,** 187-208.
17. Freistroffer, D.V., Pavlov, M.Y., MacDougall, J., Buckingham, R.H. and Ehrenberg, M. (1997) Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner *Embo J*, **16,** 4126-4133.
18. Grentzmann, G., Kelly, P.J., Laalami, S., Shuda, M., Firpo, M.A., Cenatiempo, Y. and Kaji, A. (1998) Release factor RF-3 GTPase activity acts in disassembly of the ribosome termination complex *Rna*, **4,** 973-983.
19. Poole, E. and Tate, W. (2000) Release factors and their role as decoding proteins: specificity and fidelity for termination of protein synthesis *Biochim Biophys Acta*, **1493,** 1-11.
20. Miller, W.A., Dinesh-Kumar, S.P. and Paul, C.P. (1995) Luteovirus gene expression *Critical Reviews in Plant Science*, **14,** 179-211.

21. Jacks, T., Madhani, H.D., Masiarz, F.R. and Varmus, H.E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region *Cell*, **55,** 447-458.
22. Jacks, T. and Varmus, H.E. (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting *Science*, **230,** 1237-1242.
23. Marczinke, B., Fisher, R., Vidakovic, M., Bloys, A.J. and Brierley, I. (1998) Secondary structure and mutational analysis of the ribosomal frameshift signal of rous sarcoma virus *J Mol Biol*, **284,** 205-225.
24. Parkin, N.T., Chamorro, M. and Varmus, H.E. (1992) Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo *J Virol*, **66,** 5147-5151.
25. Rettberg, C.C., Prere, M.F., Gesteland, R.F., Atkins, J.F. and Fayet, O. (1999) A three-way junction and constituent stem-loops as the stimulator for programmed -1 frameshifting in bacterial insertion sequence IS911 *J Mol Biol*, **286,** 1365-1378.
26. Belcourt, M.F. and Farabaugh, P.J. (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site *Cell*, **62,** 339-352.
27. Tate, W.P., Poole, E.S., Dalphin, M.E., Major, L.L., Crawford, D.J. and Mannering, S.A. (1996) The translational stop signal: codon with a context, or extended factor recognition element? *Biochimie*, **78,** 945-952.
28. Huang, W.M., Ao, S.Z., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R., Winge, D. and Fang, M. (1988) A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60 *Science*, **239,** 1005-1012.
29. Herr, A.J., Atkins, J.F. and Gesteland, R.F. (2000) Coupling of open reading frames by translational bypassing *Annu Rev Biochem*, **69,** 343-372.
30. Feng, Y.X., Copeland, T.D., Oroszlan, S., Rein, A. and Levin, J.G. (1990) Identification of amino acids inserted during suppression of UAA and UGA termination codons at the gag-pol junction of Moloney murine leukemia virus *Proc Natl Acad Sci U S A*, **87,** 8860-8863.
31. Li, G. and Rice, C.M. (1993) The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon *J Virol*, **67,** 5062-5067.
32. Skuzeski, J.M., Nichols, L.M., Gesteland, R.F. and Atkins, J.F. (1991) The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons *J Mol Biol*, **218,** 365-373.
33. Zerfass, K. and Beier, H. (1992) Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context *Nucleic Acids Res*, **20,** 5911-5918.
34. Brown, C.M., Dinesh-Kumar, S.P. and Miller, W.A. (1996) Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon *J Virol*, **70,** 5884-5892.
35. Prescott, C., Krabben, L. and Nierhaus, K. (1991) Ribosomes containing the C1054-deletion mutation in E. coli 16S rRNA act as suppressors at all three nonsense codons *Nucleic Acids Res*, **19,** 5281-5283.
36. Chernoff, Y.O., Newnam, G.P. and Liebman, S.W. (1996) The translational function of nucleotide C1054 in the small subunit rRNA is conserved throughout evolution: genetic evidence in yeast *Proc Natl Acad Sci U S A*, **93,** 2517-2522.
37. Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H. and Zuker, M. (1994) Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding *Proc Natl Acad Sci U S A*, **91,** 9218-9222.
38. Arkov, A.L. and Murgola, E.J. (1999) Ribosomal RNAs in translation termination: facts and hypotheses *Biochemistry (Mosc)*, **64,** 1354-1359.

39. Brown, C.M. and Tate, W.P. (1994) Direct recognition of mRNA stop signals by Escherichia coli polypeptide chain release factor two *J Biol Chem*, **269,** 33164-33170.
40. Atkins, J.F. and Gesteland, R.F. (2000) The twenty-first amino acid *Nature*, **407,** 463, 465.
41. Atkins, J.F. (1999) In et.al (ed.), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 637-673.
42. Buettner, C., Harney, J.W. and Berry, M.J. (1999) The Caenorhabditis elegans homologue of thioredoxin reductase contains a selenocysteine insertion sequence (SECIS) element that differs from mammalian SECIS elements but directs selenocysteine incorporation *J Biol Chem*, **274,** 21598-21602.
43. Heider, J., Baron, C. and Bock, A. (1992) Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein *Embo J*, **11,** 3759-3766.
44. Berry, M.J., Banu, L., Chen, Y.Y., Mandel, S.J., Kieffer, J.D., Harney, J.W. and Larsen, P.R. (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region *Nature*, **353,** 273-276.
45. Nasim, M.T., Jaenecke, S., Belduz, A., Kollmus, H., Flohe, L. and McCarthy, J.E. (2000) Eukaryotic selenocysteine incorporation follows a nonprocessive mechanism that competes with translational termination *J Biol Chem*, **275,** 14846-14852.
46. Low, S.C. and Berry, M.J. (1996) Knowing when not to stop: selenocysteine incorporation in eukaryotes *Trends Biochem Sci*, **21,** 203-208.
47. Marsh, M. and Helenius, A. (1989) Virus entry into animal cells *Adv Virus Res*, **36,** 107-151.
48. Wagner, E.K. and Hewlett, M.J. (1999), *Basic Virology*. Blackwell Science Inc., Malden, pp. 70-72.
49. Wagner, E.K. and Hewlett, M.J. (1999) In Inc, B. S. (ed.), *Basic Virology*. Blackwell Science Inc, Malden, pp. 271-273.
50. Wagner, E.K. and Hewlett, M.J. (1999) In Inc, B. S. (ed.), *Basic Virology*. Blackwell Science Inc, Malden, pp. 261-268.
51. Sawicki, S.G. and Sawicki, D.L. (1998) A new model for coronavirus transcription *Adv Exp Med Biol*, **440,** 215-219.
52. Kanamori, Y. and Nakashima, N. (2001) A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation *Rna*, **7,** 266-274.
53. Wilson, J.E., Pestova, T.V., Hellen, C.U. and Sarnow, P. (2000) Initiation of protein synthesis from the A site of the ribosome *Cell*, **102,** 511-520.
54. Hunt, S.L., Hsuan, J.J., Totty, N. and Jackson, R.J. (1999) unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA *Genes Dev*, **13,** 437-448.
55. Mitchell, S.A., Brown, E.C., Coldwell, M.J., Jackson, R.J. and Willis, A.E. (2001) Protein factor requirements of the Apaf-1 internal ribosome entry segment: roles of polypyrimidine tract binding protein and upstream of N-ras *Mol Cell Biol*, **21,** 3364-3374.
56. Pilipenko, E.V., Pestova, T.V., Kolupaeva, V.G., Khitrina, E.V., Poperechnaya, A.N., Agol, V.I. and Hellen, C.U. (2000) A cell cycle-dependent protein serves as a template-specific translation initiation factor *Genes Dev*, **14,** 2028-2045.
57. Vagner, S., Galy, B. and Pyronnet, S. (2001) Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites *EMBO Rep*, **2,** 893-898.
58. Burke, D. (2001) In Harrell, L. (ed.), Ann Arbor.
59. Kanehisa, M. (2000) *Post-genome Informatics*. Oxford University Press, Oxford.

60. Gerdes, H.H. and Kaether, C. (1996) Green fluorescent protein: applications in cell biology *FEBS Lett*, **389,** 44-47.
61. Grentzmann, G., Ingram, J.A., Kelly, P.J., Gesteland, R.F. and Atkins, J.F. (1998) A dual-luciferase reporter system for studying recoding signals *Rna*, **4,** 479-486.
62. Skuzeski, J.M., Nichols, L.M. and Gesteland, R.F. (1990) Analysis of leaky viral translation termination codons in vivo by transient expression of improved beta-glucuronidase vectors *Plant Mol Biol*, **15,** 65-79.
63. Feng, Y.X., Yuan, H., Rein, A. and Levin, J.G. (1992) Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the gag termination codon followed by an RNA pseudoknot *J Virol*, **66,** 5127-5132.
64. Beier, H. and Grimm, M. (2001) Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs *Nucleic Acids Res*, **29,** 4767-4782.
65. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure *J Mol Biol*, **288,** 911-940.
66. Zuker, M., Mathews, D.H. and Turner, D.H. (1999) In Barciszewski, J. and Clark, B. F. C. (eds.), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, pp. 11-43.
67. Kopelowitz, J., Hampe, C., Goldman, R., Reches, M. and Engelberg-Kulka, H. (1992) Influence of codon context on UGA suppression and readthrough *J Mol Biol*, **225,** 261-269.
68. Mottagui-Tabar, S., Bjornsson, A. and Isaksson, L.A. (1994) The second to last amino acid in the nascent peptide as a codon context determinant *Embo J*, **13,** 249-257.
69. Valle, R.P., Drugeon, G., Devignes-Morch, M.D., Legocki, A.B. and Haenni, A.L. (1992) Codon context effect in virus translational readthrough. A study in vitro of the determinants of TMV and Mo-MuLV amber suppression *FEBS Lett*, **306,** 133-139.
70. Honigman, A., Wolf, D., Yaish, S., Falk, H. and Panet, A. (1991) cis Acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus *Virology*, **183,** 313-319.
71. Feng, Y.X., Levin, J.G., Hatfield, D.L., Schaefer, T.S., Gorelick, R.J. and Rein, A. (1989) Suppression of UAA and UGA termination codons in mutant murine leukemia viruses *J Virol*, **63,** 2870-2873.
72. Zhang, S.P. (1996) Functional interaction between release factor one and P-site peptidyl-tRNA on the ribosome *Journal of Molecular Biology*, **261,** 98-107.
73. Brown, C.M., Stockwell, P.A., Trotman, C.N. and Tate, W.P. (1990) Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes *Nucleic Acids Res*, **18,** 6339-6345.
74. Wills, N.M., Gesteland, R.F. and Atkins, J.F. (1991) Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon *Proc Natl Acad Sci U S A*, **88,** 6991-6995.
75. Ayer, D. and Yarus, M. (1986) The context effect does not require a fourth base pair *Science*, **231,** 393-395.
76. Pedersen, W.T. and Curran, J.F. (1991) Effects of the nucleotide 3' to an amber codon on ribosomal selection rates of suppressor tRNA and release factor-1 *J Mol Biol*, **219,** 231-241.
77. Stormo, G.D., Schneider, T.D. and Gold, L. (1986) Quantitative analysis of the relationship between nucleotide sequence and functional activity *Nucleic Acids Res*, **14,** 6661-6679.
78. Li, Z., Stahl, G. and Farabaugh, P.J. (2001) Programmed +1 frameshifting stimulated by complementarity between a downstream mRNA sequence and an error-correcting region of rRNA *Rna*, **7,** 275-284.

79. Namy, O., Hatin, I. and Rousset, J.P. (2001) Impact of the six nucleotides downstream of the stop codon on translation termination *EMBO Rep*, **2,** 787-793.
80. Mottagui-Tabar, S., Tuite, M.F. and Isaksson, L.A. (1998) The influence of 5' codon context on translation termination in Saccharomyces cerevisiae *Eur J Biochem*, **257,** 249-254.
81. Li, G.P. and Rice, C.M. (1989) Mutagenesis of the in-frame opal termination codon preceding nsP4 of Sindbis virus: studies of translational readthrough and its effect on virus replication *J Virol*, **63,** 1326-1337.
82. Zhang, S., Ryden-Aulin, M. and Isaksson, L.A. (1999) Interaction between a mutant release factor one and P-site peptidyl-tRNA is influenced by the identity of the two bases downstream of the stop codon UAG *FEBS Lett*, **455,** 355-358.
83. Jamjoom, G.A., Naso, R.B. and Arlinghaus, R.B. (1977) Further characterization of intracellular precursor polyproteins of Rauscher leukemia virus *Virology*, **78,** 11-34.
84. Phillips-Jones, M.K., Hill, L.S., Atkinson, J. and Martin, R. (1995) Context effects on misreading and suppression at UAG codons in human cells *Mol Cell Biol*, **15,** 6593-6600.
85. Urban, C., Zerfass, K., Fingerhut, C. and Beier, H. (1996) UGA suppression by tRNACmCATrp occurs in diverse virus RNAs due to a limited influence of the codon context *Nucleic Acids Res*, **24,** 3424-3430.

VITA 2

Lance Harrell

Candidate for the Degree of

Master of Science

Thesis: PREDOMINACE OF SIX HEXANUCLEOTIDE RECODING SIGNALS 3' OF READ-THROUGH STOP CODONS

Major Field: Biochemistry and Molecular Biology

Biographical:

Education: Graduated from Pequot Lakes High School, Pequot Lakes, Minnesota in May 1993; received Bachelor of Arts degree in Biology from Concordia College, Moorhead, Minnesota. Completed the requirements for the Master of Science degree with a major in Biochemistry and Molecular Biology at Oklahoma State University in May 2002.

Experience: Employed by Oklahoma State University, Department of Biochemistry and Molecular Biology as a graduate research assistant; Oklahoma State University, Department of Biochemistry and Molecular Biology, 1999 to present.